

GenCore version 5.1.6
Copyright (C) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 21:41:43 ; Search time 2063 Seconds
(without alignments)
2721.442 Million cell updates/sec

Title: US-09-673-716-1
Perfect score: 231
Sequence: 1 gtggggtcgcggactaagca.....ttcaccatgaggtgtgctt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_fod.*
26: em_gss_bhg.*
27: em_gss_vri.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	231	100.0	477	28	AZ231159	AZ231159 RPCI-23-5
2	231	100.0	501	28	AZ440299	AZ440299 1M0231K17
3	231	100.0	527	10	BG087407	BG087407 H3139B12-
4	231	100.0	558	28	BH036520	BH036520 RPCI-24-2

5	231	100.0	559	28	BH105759	BH105759 RPCI-24-2
6	231	100.0	577	28	AZ420721	AZ420721 1M0198L07
7	231	100.0	587	28	AZ832364	AZ832364 2M0112B13
8	231	100.0	627	28	AZ495883	AZ495883 1M0332E03
9	231	100.0	647	28	AZ993990	AZ993990 2M0279D16
10	231	100.0	706	14	CB056516	CB056516 NISC-117
11	231	100.0	721	28	AZ994818	AZ994818 2M0280J24
12	229.4	99.3	472	28	AZ310386	AZ310386 1M0025O13
13	229.4	99.3	505	28	AZ054649	AZ054649 RPCI-23-4
14	229.4	99.3	518	9	AW558986	AW558986 L0302F10-
15	229.4	99.3	520	12	BM219777	BM219777 C0929H06-
16	229.4	99.3	534	28	AZ341190	AZ341190 1M0073M24
17	229.4	99.3	535	28	AZ261410	AZ261410 1M0048D11
18	229.4	99.3	539	28	AZ965228	AZ965228 2M0235A17
19	229.4	99.3	557	9	AW554086	AW554086 L0235F05-
20	229.4	99.3	558	28	AZ987016	AZ987016 2M0269M23
21	229.4	99.3	584	28	AZ432413	AZ432413 1M0217N15
22	229.4	99.3	588	28	AZ242504	AZ242504 RPCI-23-8
23	229.4	99.3	662	28	AZ333042	AZ333042 1M0061J16
24	229.4	99.3	678	14	BY731717	BY731717 BY731717
25	229.4	99.3	686	28	BH092157	BH092157 RPCI-24-2
26	229.4	99.3	707	28	BH117452	BH117452 RPCI-24-2
27	229.4	99.3	834	28	BH062187	BH062187 RPCI-24-3
28	229.4	99.3	4484	11	AK033082	AK033082 Mus_muscu
29	228.4	98.9	606	12	BI134544	BI134544 UI-M-BH3
30	227.8	98.6	309	28	BH068706	BH068706 RPCI-24-3
31	227.8	98.6	457	28	AZ071850	AZ071850 RPCI-23-4
32	227.8	98.6	492	9	AW552509	AW552509 L0213E08-
33	227.8	98.6	531	28	AZ636367	AZ636367 1M0495J10
34	227.8	98.6	557	28	BH078165	BH078165 RPCI-24-2
35	227.8	98.6	559	13	BQ553681	BQ553681 H4033F02-
36	227.8	98.6	596	28	AZ830822	AZ830822 2M0110H14
37	227.8	98.6	646	28	BH099567	BH099567 RPCI-24-3
38	227.8	98.6	678	28	AZ897416	AZ897416 RPCI-24-2
39	227.8	98.6	686	28	AZ259805	AZ259805 RPCI-23-1
40	227.8	98.6	706	28	AZ797045	AZ797045 2M0053J06
41	227.8	98.6	793	28	AZ719135	AZ719135 RPCI-24-1
42	227.8	98.6	838	28	BH027115	BH027115 RPCI-24-2
43	227.4	98.4	553	28	AZ792711	AZ792711 2M0045C19
44	227.4	98.4	642	28	AZ090071	AZ090071 RPCI-23-4
45	227.4	98.4	675	28	AZ957637	AZ957637 2M0224N06

ALIGNMENTS

RESULT 1
AZ231159
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ231159
RPCI-23-51K23-TJ RPCI-23
genomic survey sequence.
GI:8539205
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 477)
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aklnrst
, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
Other GSSs: RPCI-23-51K23-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
Plate: 51 row: K column: 23
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .477
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-51K23"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 94 a 110 c 143 g 130 t
ORIGIN
Query Match 100.0%; Score 231; DB 28; Length 477;
Best Local Similarity 100.0%; Pred. No. 1.5e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGTGCGAGCTAAGCACTACACAGAGGATAGCTTGTGTGGCATCTCTGGAAGG 60
|||||
Db 232 GTGGGTGCGAGCTAAGCACTACACAGAGGATAGCTTGTGTGGCATCTCTGGAAGG 291
QY 61 CAGCTCTGATGATGAGTTGAGTTCAGTGTCTAGTCTCCCTCCCGAGGAAAACGACAG 120
|||||
Db 292 CAGCTCTGATGATGAGTTGAGTTCAGTGTCTAGTCTCCCTCCCGAGGAAAACGACAG 351
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAGGGATGTTTGTGAGGGCCCT 180
|||||
Db 352 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAGGGATGTTTGTGAGGGCCCT 411
QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
|||||
Db 412 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 462

RESULT 2
AZ440299 501 bp DNA linear GSS 03-OCT-2000
LOCUS 1M0231K17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0231K17 F, genomic survey sequence.
ACCESSION AZ440299
VERSION AZ440299.1 GI:10564312
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhauser,A.
and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT

84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0231 row: K column: 17
Seq primer: CGTTGTAACAGAGCGCCAGT
Class: plasmid ends
High quality sequence stop: 501.
Location/Qualifiers
1. .501
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0231K17"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWB42 (gill4732114[gb]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES
source

Location/Qualifiers
1. .501
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0231K17"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWB42 (gill4732114[gb]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 119 a 123 c 136 g 123 t
ORIGIN
Query Match 100.0%; Score 231; DB 28; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.5e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGTGCGAGCTAAGCACTACACAGAGGATAGCTTGTGTGGCATCTCTGGAAGG 60
|||||
Db 94 GTGGGTGCGAGCTAAGCACTACACAGAGGATAGCTTGTGTGGCATCTCTGGAAGG 153
QY 61 CAGCTCTGATGATGAGTTGAGTTCAGTGTCTAGTCTCCCTCCCGAGGAAAACGACAG 120
|||||
Db 154 CAGCTCTGATGATGAGTTGAGTTCAGTGTCTAGTCTCCCTCCCGAGGAAAACGACAG 213
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAGGGATGTTTGTGAGGGCCCT 180
|||||
Db 214 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTCAGGGATGTTTGTGAGGGCCCT 273
QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
|||||
Db 274 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 324

RESULT 3
BG087407 527 bp mRNA linear EST 13-JUN-2003
LOCUS H3139B12-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
DEFINITION H3139B12 5', mRNA sequence.
ACCESSION BG087407
VERSION BG087407.1 GI:12569971
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Qy	1	GTGGGTGCAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGTGAAGG	60
Db	376	GTGGGTGCCAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGTGAAGG	317
Qy	61	CAGCTCTGATTGCATGAAGGTTCAAGTCTCCTAGTTCCTTCCCCCAGAAAAACGACACG	120
Db	316	CAGCTCTGATTGCATGAAGGTTCAAGTCTCCTAGTTCCTTCCCCCAGAAAAACGACACG	257
Qy	121	GGAGCTGGCCAGACCTCTCTGGTGATAGCCCTAAGGGATGGTTTTGTGTAGGGGCCCT	180
Db	256	GGAGCTGGCCAGACCTCTCTGGTGATAGCCCTAAGGGATGGTTTTGTGTAGGGGCCCT	197

```

QY 181 ATGCTTCACACATGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 196 ATGCTTCACACATGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 146

RESULT 5
BH105759
LOCUS BH105759 559 bp DNA linear GSS 19-JUL-2001
DEFINITION RPCI-24-237C16.TJ RPCI-24 Mus musculus genomic clone RPCI-24-237C16
, genomic survey sequence.
ACCESSION BH105759
VERSION BH105759.1 GI:14935045
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 559)
AUTHORS Zhao S., Nieman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,
Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E.,
Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end
page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
Plate: 237 row: C column: 16
Seq primer: SP6
Class: BAC ends.
FEATURES             location/Qualifiers
   source
1..559
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-237C16"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notes="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT 119 a 132 c 159 g 149 t
ORIGIN
Query Match 100.0%; Score 231; DB 28; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.6e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGCAGCTAAGCACTCACAGAGTAGCTTGCTGTGGCATCCTCTGGGAGC 60
Db 243 GTGGGTGGCAGCTAAGCACTCACAGAGTAGCTTGCTGTGGCATCCTCTGGGAGC 302

QY 61 CAGCTCTGATTGATGAAGTTTCAGTGTCTAGTTCCTCCCTCCAGGAAACACACACG 120
Db 303 CAGCTCTGATTGATGAAGTTTCAGTGTCTAGTTCCTCCCTCCAGGAAACACACACG 362

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 180
Db 363 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 422

QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231

```

```

Db 423 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 473

RESULT 6
AZ420721/c
LOCUS AZ420721 577 bp DNA linear GSS 03-OCT-2000
DEFINITION IM0198L07R Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0198L07 R, genomic survey sequence.
ACCESSION AZ420721
VERSION AZ420721.1 GI:10544734
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 577)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0198 row: L column: 07
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 577.
FEATURES             location/Qualifiers
   source
1..577
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCIM0198L07"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGCIM library"
/notes="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gii4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT 152 a 171 c 131 g 123 t
ORIGIN
Query Match 100.0%; Score 231; DB 28; Length 577;
Best Local Similarity 100.0%; Pred. No. 1.6e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCTGTGGAAGG 60
 |||||||
 Db 374 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCTGTGGAAGG 315
 |||||||
 QY 61 CAGCTCTGATGTCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAACACGACG 120
 |||||||
 Db 314 CAGCTCTGATGTCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAACACGACG 255
 |||||||
 QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180
 |||||||
 Db 254 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 195
 |||||||
 QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTTGCTT 231
 |||||||
 Db 194 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTTGCTT 144
 |||||||

RESULT 7
 AZ832364/c
 LOCUS 587 bp DNA linear GSS 20-FEB-2001
 DEFINITION 2M0112B13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC2M0112B13 R, genomic survey sequence.
 AZ832364
 VERSION
 ACCESSION AZ832364.1 GI:13002272
 KEYWORDS
 SOURCE
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 587)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0112 row: B column: 13
 Seq primer: CACACAGGAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 587.
 Location/Qualifiers
 1. .587
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0112B13"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PWD42 (g114732114|gb|AF129072.1), a copy-number

Inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adapted mouse DNA was annealed to
 adapted vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 159 a 168 c 135 g 125 t
 ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 587;
 Best Local Similarity 100.0%; Pred. No. 1.6e-60;
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCTGTGGAAGG 60
 |||||||
 Db 246 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTTGGCATCCTGTGGAAGG 187
 |||||||

QY 61 CAGCTCTGATGTCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAACACGACG 120
 |||||||
 Db 186 CAGCTCTGATGTCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAACACGACG 127
 |||||||

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 180
 |||||||
 Db 126 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGGTTTGTAGGGCCCT 67
 |||||||

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTTGCTT 231
 |||||||
 Db 66 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCCACCATGAGGCTTGCTT 16
 |||||||

RESULT 8
 AZ495883/c
 LOCUS 627 bp DNA linear GSS 05-OCT-2000
 DEFINITION 1M0332E03F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0332E03 F, genomic survey sequence.
 AZ495883
 VERSION
 ACCESSION AZ495883.1 GI:10671633
 KEYWORDS
 SOURCE
 ORGANISM Mus musculus (house mouse)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 627)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
 ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,
 and Wright,D., Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0332 row: E column: 03
 Seq primer: CGTTGTAAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 627.
 Location/Qualifiers
 1. .627
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0332E03"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES
 source

/note="vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 160 a 190 c 142 g 135 t
ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.7e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGGAAGG 60
Db 441 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGGAAGG 382

Qy 61 CACGCTCGATGATGAGGTTGAGTCTCAGTCTCAGTCTCCTCCCTCCCGAGGAAACGACACG 120
Db 381 CACGCTCGATGATGAGGTTGAGTCTCAGTCTCAGTCTCCTCCCTCCCGAGGAAACGACACG 322

Qy 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
Db 321 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 262

Qy 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 261 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 211

RESULT 9
AZ993990 647 bp DNA linear GSS 27-APR-2001
LOCUS 2M0279D16F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
DEFINITION clone UUGC2M0279D16 F, genomic survey sequence.
ACCESSION AZ993990
VERSION AZ993990.1 GI:13865217
KEYWORDS GSS.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 647)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0279 row: D column: 16

Seq primer: CGTTGTAAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 647.
Location/Qualifiers
source 1..647
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0279D16"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/note="vector: PWD42nv; Purified genomic DNA from M. Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gii4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 149 a 146 c 185 g 167 t
ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 647;
Best Local Similarity 100.0%; Pred. No. 1.7e-60;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGGAAGG 60
Db 235 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGGCATCTCTGGAAGG 294

Qy 61 CACGCTCGATGATGAGGTTGAGTCTCAGTCTCAGTCTCCTCCCTCCCGAGGAAACGACACG 120
Db 295 CACGCTCGATGATGAGGTTGAGTCTCAGTCTCAGTCTCCTCCCTCCCGAGGAAACGACACG 354

Qy 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
Db 355 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 414

Qy 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 415 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 465

RESULT 10
CB056516/c 706 bp mRNA linear EST 17-JAN-2003
LOCUS NISC-jj37f07.w1 Soares NMBP13-15 Mus musculus cDNA clone
DEFINITION IMAGE:4848516 5', mRNA sequence.
ACCESSION CB056516
VERSION CB056516.1 GI:27794803
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 706)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: c9abps@email.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM10779 row: K column: 13
Seq primer: T7 primer.

FEATURES

source

Location/Qualifiers

1..706

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="IMAGE:4848516"

/tissue_type="pituitary gland"

/dev_stage="juvenile, 13-15 days"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pT7T3D-PacI; Site_1: NotI;

Site_2: EcoRI; 1st strand cDNA was primed with a NotI-
oligo(dT) primer

5'-ACTGCAAGATTTCGGCGCGCTGACGATCTTTTCTTTTCTTTT-3';
double-stranded cDNA was ligated to EcoRI adaptors

5'-AATTCGCGCAGAG-3' and 5'-CTCGTCCG-3' (Pharmacia),
digested with NotI and cloned into the NotI and EcoRI
sites of the pT7T3D-PacI vector. Library went through one
round of normalization, and was constructed in the
laboratory of M. Bento Soares (University of Iowa)."

157 a 199 c 172 g 178 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 231; DB 14; Length 706;

Best Local Similarity 100.0%; Pred. No. 1.8e-60;

Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCCTGTGGAAGG 60

Db 634 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCCTGTGGAAGG 575

QY 61 CAGCTCTGATTCGATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAACGACAG 120

Db 574 CAGCTCTGATTCGATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAACGACAG 515

QY 121 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 180

Db 514 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 455

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231

Db 454 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 404

RESULT 11

AZ994818/c

LOCUS

DEFINITION 721 bp DNA linear GSS 27-APR-2001
clone UUGC2M0280J24 F, genomic survey sequence.

ACCESSION AZ994818

VERSION AZ994818.1 GI:13866045

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 721)

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly

,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.

and Wright,D.,Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb

JOURNAL

COMMENT

plasmid inserts

Unpublished

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0280 row: J column: 24

Seq primer: CGTTGTAACACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 721.

FEATURES

source

Location/Qualifiers

1..721

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0280J24"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, Ti-resistant, P-"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (G14732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

191 a 202 c 166 g 162 t

BASE COUNT

ORIGIN

Query Match 100.0%; Score 231; DB 28; Length 721;

Best Local Similarity 100.0%; Pred. No. 1.8e-60;

Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCCTGTGGAAGG 60

Db 300 GTGGGTGGGAGGCTAAGCACTGCACAGAGGATAGTCTGCTGGCATCCTGTGGAAGG 241

QY 61 CAGCTCTGATTCGATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAACGACAG 120

Db 240 CAGCTCTGATTCGATCAAGGTCAGTCTCTAGTTCCTTCCCTCCCGAGGAAACGACAG 181

QY 121 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 180

Db 180 GGAGCTGGCCACAGACCTCTCTGGGTGATGAGCCTAAGGATGTTTGTGTAGGGCCCT 121

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231

Db 120 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 70

RESULT 12

AZ310386/c

LOCUS

DEFINITION 472 bp DNA linear GSS 29-SEP-2000
clone UUGC1M0025013 F, genomic survey sequence.

A2310386
 VERSION A2310386.1 GI:10352324
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0025 row: O column: 13
 Seq primer: CGGTGTAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 472.

FEATURES
 Source
 1..472

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGGCLM0025013"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UGGCLM library"
 /note="vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydronically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (gll4732114(gb)AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 133 a 140 c 104 g 95 t
 ORIGIN
 Query Match 99.3%; Score 229.4; DB 28; Length 472;
 Best Local Similarity 99.6%; Pred. No. 4.6e-60;
 Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTCATCTGTGGAAGG 60
 |||||
 281 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTCATCTGTGGAAGG 222
 |||||
 QY 61 CAGCTCTGATTCATGAAGTTTCAGTGTCTTCCCTTCCCGGAGGAAAAACGACACG 120
 |||||
 221 CATGCTGATTCATGAAGTTTCAGTGTCTTCCCTTCCCGGAGGAAAAACGACACG 162
 |||||
 QY 121 GGAGCTGGCCAAAGCACTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGAGGGCCCT 180
 |||||

Db 161 GGAGCTGGCCAAAGCACTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGAGGCCCT 102
 |||||
 QY 181 ATGCTTGCACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
 |||||
 Db 101 ATGCTTGCACACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 51
 |||||
 RESULT 13
 LOCUS AZ054649 505 bp DNA linear GSS 30-MAR-2000
 DEFINITION RPCI-23-401C1.TJ RPCI-23 Mus musculus genomic clone RPCI-23-401C1,
 genomic survey sequence.
 ACCESSION AZ054649
 VERSION AZ054649.1 GI:7345885
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P., and Fraser,C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished
 COMMENT Other GSSs: RPCI-23-401C1.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 401 row: C column: 1
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 1..505

/organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-401C1"
 /sex="Female"
 /lab_host="DH10B"
 /clone_lib="RPCI-23"
 /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 100 a 114 c 147 g 144 t
 ORIGIN

Query Match 99.3%; Score 229.4; DB 28; Length 505;
 Best Local Similarity 99.6%; Pred. No. 4.8e-60;
 Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTCATCTGTGGAAGG 60
 |||||
 233 GTGGGTGCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTCATCTGTGGAAGG 292
 |||||
 QY 61 CAGCTCTGATTCATGAAGTTTCAGTGTCTTCCCTTCCCGGAGGAAAAACGACACG 120
 |||||


```

Db      293 CATGCTGATGTCATGAGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 352
QY      121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGTAGGGCCCT 180
Db      353 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGTAGGGCCCT 412
QY      181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db      413 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 463

RESULT 14
AW558986/c
LOCUS      518 bp mRNA linear EST 31-AUG-2000
DEFINITION clone L0302F10 3', mRNA sequence.
ACCESSION AW558986
VERSION   AW558986
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE
AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
          M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
          III, Becker,K.G. and Ko,M.S.H.
          1 (bases 1 to 518)
TITLE     Genome-wide expression profiling of mid-gestation placenta and
          embryo using a 15,000 mouse developmental cDNA microarray
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
MEDLINE   20381348
PUBMED    10922068
COMMENT   Contact: George J. Kargul
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
          Email: cdna@lgsun.grc.nia.nih.gov
          Plate: L0302 row: F column: 10
          Seq primer: -21M13 Forward
          High quality sequence stop: 518
          POLYA=Yes.

FEATURES
         Location/Qualifiers
            1..518
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="niaEST:L0302F10-3"
               /db_xref="taxon:10090"
               /clone="L0302F10"
               /sex="female"
               /dev_stage="Newborn Ovary"
               /lab_host="DH10B"
               /clone_lib="NIA Mouse Newborn Ovary cDNA Library"
               /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
               Site.1: SalI; Site.2: NotI; Total RNAs were extracted from
               7 Newborn Ovary. The double-stranded cDNA was synthesized
               by Gibco's kit with an Oligo(dT) primer [NotI
               primer-adaptor from GibcoBRL]
               [5'-pGACTAGTCTAGATCGCGAGCGCGCTTTT-3'] from
               2.5ug of total RNA. The double-stranded cDNAs were
               treated with T4 DNA polymerase and purified by
               ethanol-precipitation. The cDNAs were ligated to
               lone-linker LL-Sal3 (include SalI sequence). The cDNAs
               were purified by phenol/chloroform and separated from
               free linkers by Centricon 100. Then, cDNAs were amplified
               by long-range high fidelity PCR using Takara's Ex Taq
               polymerase. Then, the cDNAs were purified by
               phenol/chloroform and by Centricon 100. The cDNAs were
               digested with SalI and NotI enzymes. Then, the cDNAs were
               size selected by Gibco's Size Fractionation Column. The
               cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid
               vector. The DH10B E. coli host was transformed with the
               ligation mixture by chemical method. The library was

```

```

constructed by Xiaohong Wang and Yulan Piao.*
BASE COUNT 126 a 138 c 131 g 123 t
ORIGIN
Query Match 99.3%; Score 229.4; DB 9; Length 518;
Best Local Similarity 99.6%; Pred. No. 4.9e-60;
Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATGAGTTCGTGTGTCATCCTGTGGAAGG 60
Db 465 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATGAGTTCGTGTGTCATCCTGTGGAAGG 406
QY 61 CAGCTCTGATTCGATCAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 120
Db 405 CAGCTCTGATTCGATCAAGGTTTCAGTGTCTCTAGTTCCTCCCTCCAGGAAACGACACG 346
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGTAGGGCCCT 180
Db 345 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTGTAGGGCCCT 286
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 285 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 235

RESULT 15
BM219777/c
LOCUS      520 bp mRNA linear EST 07-JUN-2003
DEFINITION C0929H06-3 NIA Mouse 12.5-dpc Male Genital Ridge/Mesonephros cDNA
          Library (Long) Mus musculus cDNA clone NIA: C0929H06 IMAGE:30037433
          3', mRNA sequence.
ACCESSION BM219777
VERSION   BM219777.2 GI:31478475
KEYWORDS  EST.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
REFERENCE
AUTHORS   Piao,Y., Ko,M.S.H., Lim,M.K. and Ko,M.S.H.
          Construction of long-transcript enriched cDNA libraries from
          submicrogram amounts of total RNAs by a universal PCR amplification
          method
JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE   21429098
PUBMED    11544199
COMMENT   On Dec 14, 2001 this sequence version replaced gi:17779702.
          Contact: Dawood B. Dudekula
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
          Email: cdna@lgsun.grc.nia.nih.gov
          Plate: C0929 row: H column: 06
          Seq primer: -21M13 Forward
          High quality sequence stop: 520
          POLYA=Yes.

FEATURES
         Location/Qualifiers
            1..520
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="niaEST:C0929H06-3"
               /db_xref="taxon:10090"
               /clone="NIA:C0929H06 IMAGE:30037433"
               /sex="Male"
               /tissue_type="Male genital ridge/mesonephros"
               /dev_stage="12.5-dpc"
               /lab_host="DH10B"
               /clone_lib="NIA Mouse 12.5-dpc Male Genital
               Ridge/Mesonephros cDNA Library (Long)"
               /note="Vector: pSPORT1 (Invitrogen); Site.1: SalI; Site.2:
               NotI; Mouse cDNA project by the Laboratory of Genetics,
               National Institute on Aging (NIA), Intramural Research

```

Program, NIH (<http://lsun.grc.nia.nih.gov/cdna>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pgactagtcttagatcggagcgccgcttttttttttt-3'] from 1.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.4 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 122 a 140 c 134 g 124 t
ORIGIN

Query Match 99.3%; Score 229.4; DB 12; Length 520;
Best Local Similarity 99.6%; Pred. No. 4.9e-60;
Matches 230; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTGGGTGCGAGGTAAGCACTGCACAGAGATAGCTTGCTGTGGCATCCTGTGGAAGG 60
 |||||
Db 465 GTGGGTGCGAGGTAAGCACTGCACAGAGATAGCTTGCTGTGGCATCCTGTGGAAGG 406
 |||||

Qy 61 CAGCTCTGATGTCATGAAGTTTCAGTGTCCCTTCCTCCCGGAGGAGGACG 120
 |||||
Db 405 CAGCTCTGATGTCATGAAGTTTCAGTGTCCCTTCCTCCCGGAGGAGGACG 346
 |||||

Qy 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
 |||||
Db 345 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 286
 |||||

Qy 181 ATGCTTGACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
 |||||
Db 285 ATGCTTGACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 235
 |||||

Search completed: October 9, 2003, 22:40:58
Job time : 2067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:44:04 ; Search time 206 Seconds
(without alignments)
2909.716 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 ggggggctggagctaaagca.....ttcaccatgaggtgtgtt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2.6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2.6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2.6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2.6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2.6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	214.2	92.7	185548	14	US-10-175-523-62
C 2	92.6	40.1	432	10	US-09-933-797-502
C 3	89.4	38.7	1913	12	US-10-225-810-40
C 4	86.2	37.3	1434	10	US-09-799-462A-13
C 5	86.2	37.3	1434	11	US-09-836-911A-13
C 6	86.2	37.3	1434	13	US-10-125-767-13
C 7	86.2	37.3	1434	14	US-10-151-081-13
C 8	86.2	37.3	1434	14	US-10-287-313-13
C 9	86.2	37.3	1434	14	US-10-219-694-13
C 10	81.4	35.2	1400	10	US-09-799-462A-14
C 11	81.4	35.2	1400	11	US-09-836-911A-14
C 12	81.4	35.2	1400	13	US-10-125-767-14
C 13	81.4	35.2	1400	14	US-10-151-081-14
C 14	81.4	35.2	1400	14	US-10-287-313-14
C 15	81.4	35.2	1400	14	US-10-219-694-14
C 16	70.4	30.5	251364	14	US-10-175-523-62

Sequence 61, Appl	US-10-175-523-61	14	251364	30.5	70.4	17
Sequence 79, Appl	US-10-175-523-79	14	251364	30.5	70.4	18
Sequence 36, Appl	US-09-563-728A-36	11	147	14.7	34	19
Sequence 4575, Ap	US-09-867-701-4575	10	475	13.2	30.4	20
Sequence 39036, A	US-10-027-632-39036	13	447	13.0	30	21
Sequence 65139, A	US-10-027-632-65139	13	447	13.0	30	22
Sequence 2639, Ap	US-09-918-995-2639	11	457	13.0	30	23
Sequence 112, App	US-10-231-417-112	12	609	13.0	30	24
Sequence 151950,	US-10-027-632-151950	13	968	12.8	29.6	25
Sequence 151951,	US-10-027-632-151951	13	968	12.8	29.6	26
Sequence 151952,	US-10-027-632-151952	13	968	12.8	29.6	27
Sequence 181112,	US-10-027-632-181112	13	499	12.7	29.4	28
Sequence 109, App	US-10-102-627-109	14	32187	12.7	29.4	29
Sequence 95, Appl	US-09-250-611-95	10	540	12.6	29.2	30
Sequence 112118,	US-10-027-632-112118	13	2977	12.6	29.2	31
Sequence 112119,	US-10-027-632-112119	13	2977	12.6	29.2	32
Sequence 10697, A	US-09-783-590-10697	10	360	12.4	28.6	33
Sequence 1096, Ap	US-09-764-847-1096	10	427	12.4	28.6	34
Sequence 1096, Ap	US-10-092-154-1096	14	427	12.4	28.6	35
Sequence 1096, Ap	US-10-027-632-70833	13	564	12.4	28.6	36
Sequence 70834, A	US-10-027-632-70834	13	564	12.4	28.6	37
Sequence 312268,	US-10-027-632-312268	13	564	12.4	28.6	38
Sequence 312269,	US-10-027-632-312269	13	564	12.4	28.6	39
Sequence 5, Appl	US-10-080-381B-5	12	1197	12.4	28.6	40
Sequence 45, Appl	US-10-240-985-45	12	6549	12.4	28.6	41
Sequence 102751,	US-10-027-632-102751	13	2449	12.3	28.4	42
Sequence 112101,	US-10-027-632-112101	13	2449	12.3	28.4	43
Sequence 1615, Ap	US-09-764-847-1615	10	10472	12.3	28.4	44
Sequence 1615, Ap	US-10-092-154-1615	14	10472	12.3	28.4	45

ALIGNMENTS

RESULT 1

US-10-175-523-62/c
; Sequence 62, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Priithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/1J795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; PRIOR FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62
; LENGTH: 185548
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-62

Query Match 92.7%; Score 214.2; DB 14; Length 185548;
Best Local Similarity 98.3%; Pred. No. 1.3e-67;
Matches 227; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 GTGGGTCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGTGGAAGG 60
Db 52312 GTGGGTCGAGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGTGGAAGG 52253
QY 61 CACGTCTGATTGCAATGAAGTTCACTGCTAGTTCCTTCCCGCCAGAAAACGACACG 120
Db 52252 TAGTCTGCA-TCCATGAAGGTTCACTGCTAGTTCCTTCCCGCCAGAAAACGACACG 52194
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTGTGTAGGGCCCT 180
Db 52193 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGGATGTTTGTGTAGGGCCCT 52134
QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCCATCAGAGCTTGTCTT 231
Db 52133 ATGCTTGCACATGGGGATCAGACCTCTACCTTACCCCATCAGAGCTTGTCTT 52083

RESULT 2

US-09-933-797-502/c
; Sequence 502, Application US/09933797
; Patent No. US20020155119A1
; GENERAL INFORMATION:
; APPLICANT: Robert A. Sikes et al.
; TITLE OF INVENTION: Isolation and Use of Fetal Urogenital
; FILE REFERENCE: 9901-007-999
; CURRENT APPLICATION NUMBER: US/09/933,797
; CURRENT FILING DATE: 2001-08-22
; PRIOR FILING DATE: 2000-01-14
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: PCT/US99/10746
; PRIOR FILING DATE: 1999-05/14
; PRIOR APPLICATION NUMBER: 60/085,383
; PRIOR FILING DATE: 1998-05-14
; NUMBER OF SEQ ID NOS: 811
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 502
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Murine
US-09-933-797-502

Query Match 40.1%; Score 92.6; DB 10; Length 432;
Best Local Similarity 76.0%; Pred. No. 9.9e-24;
Matches 130; Conservative 0; Mismatches 34; Indels 7; Gaps 1;
QY 55 GGAAGGCACCTCTGATTGTCATGAAGTTTCAGTGTCTTCCCTTCCCGCCAGAAAAC 114
Db 416 GGGAGACATGTCATCTTTCATGAGGTTTCAGTGTCTTCCCTTCCCGCCAGAAAAC 357
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 356 GACACGGGAGCAGTCAAGGTTGCTCTGGGTAAAAGCCTGAGCCTTAAGAGCTAATCCT 297
QY 168 GTGATGGCCCTTACCTTGCACACGAGGATGAGCCTTACCTTACCTCACC 218
Db 296 GTACATGGCTCTTACCTACACACTGGGGATTGACCTTATCTCCACTC 246

RESULT 3

US-10-225-810-40/c
; Sequence 40, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Birmingham, Jr., John R.
; TITLE OF INVENTION: Transdorins and Methods of Using Transdorin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40

; LENGTH: 1913
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-225-810-40
Query Match 38.7%; Score 89.4; DB 12; Length 1913;
Best Local Similarity 74.9%; Pred. No. 2.2e-22;
Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;
QY 55 GGAAGGCACCTCTGATTGTCATGAAGTTTCAGTGTCTTCCCTTCCCGCCAGAAAAC 114
Db 1347 GGGAGACATGTCATCTTTCATGAAGTTTCAGTGTCTTCCCTTCCCGCCAGAAAAC 1288
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 1287 GACACGGGAGCAGTCAAGGTTGCTCTGGGTAAAAGCCTTAAGCCTTAAGAGCTAATCCT 1228
QY 168 GTGATGGCCCTTACCTTGCACACGAGGATGAGCCTTACCTTACCTCACC 218
Db 1227 GTACATGGCTCTTACCTACACACTGGGGATTGACCTTATCTCCACTC 1177

RESULT 4

US-09-799-462A-13
; Sequence 13, Application US/09799462A
; Patent No. US20020160970A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,462A
; FILING DATE: 10-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO

; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-799-462A-13

Query Match 37.3%; Score 86.2; DB 10; Length 1434;
Best Local Similarity 73.7%; Pred. No. 3.1e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAC 114
Db 900 GGGGACATGTCATCTTTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAC 959
QY 115 GACAGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGATGTTT 167
Db 960 GACAGGGAGCTGGCCAGGTTGCTCTGGGTAAGGCTGTGAGCTGGGAGCTAATCT 1019
QY 168 GTGAGGGCCCTATGCTTGACACTGGGGATTCACCTTACCTTCACCC 218
Db 1020 GTACATGGCTCCTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 1070

RESULT 5

US-09-836-911A-13
; Sequence 13, Application US/09836911A
; Publication No. US20030033617A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula

; Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES

; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4350 La Jolla Village Drive, 6th Floor
; CITY: San Diego
; STATE: CA

; COUNTRY: USA
; ZIP: 92122

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/836,911A
; FILING DATE: 17-Apr-2002

; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996

ATTORNEY/AGENT INFORMATION:

; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-4021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-09-836-911A-13

Query Match 37.3%; Score 86.2; DB 11; Length 1434;
Best Local Similarity 73.7%; Pred. No. 3.1e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAC 114
Db 900 GGGGACATGTCATCTTTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCTCCAGGAAAC 959
QY 115 GACAGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGATGTTT 167
Db 960 GACAGGGAGCTGGCCAGGTTGCTCTGGGTAAGGCTGTGAGCTGGGAGCTAATCT 1019
QY 168 GTGAGGGCCCTATGCTTGACACTGGGGATTCACCTTACCTTCACCC 218
Db 1020 GTACATGGCTCCTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 1070

RESULT 6

US-10-125-767-13
; Sequence 13, Application US/10125767
; Publication No. US20020160410A1
; GENERAL INFORMATION:

; APPLICANT: Hadlaczky, Gyula

; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
; METHODS
; FOR PREPARING ARTIFICIAL CHROMOSOMES

; NUMBER OF SEQUENCES: 34

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Heller Ehrman White & McAuliffe LLP
; STREET: 4350 La Jolla Village Drive, 7th Floor
; CITY: San Diego

; STATE: CA

; COUNTRY: USA

; ZIP: 92122

COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/125,767
; FILING DATE: 17-Apr-2002

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/724,693
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996

ATTORNEY/AGENT INFORMATION:

; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-4021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-287-313-13

Query Match 37.3%; Score 86.2; DB 14; Length 1434;
Best Local Similarity 73.7%; Pred. No. 3.1e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;
QY 55 GGAAGCAGCTCTGATTGTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGGAGAAAAC 114
DB 900 GGGGACATGTCTATCTTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGGAGAAAAC 959
QY 115 GACACGGGAGCTGCGCCAAAGCTCTCTGGTGA-----TGAGCCTAAGGGATGTTT 167
DB 960 GACACGGGAGCAGTCAAGGTTCTCTGGTAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019
QY 168 GTGATGGGCCCTTATCTTGCACACATGGGATCAGACCTCTACCTTCAACC 218
DB 1020 GTACATGGCTCTTACCTACACATGGGATTTGACCTCTATCTCCACTC 1070

RESULT 9

US-10-219-694-13
; Sequence 13, Application US/10219694
; Publication No. US20030108914A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS
; FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP
; STREET: 4350 La Jolla Village Drive, 7th Floor
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/219,694
; FILING DATE: 14-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779

; REFERENCE/DOCKET NUMBER: 24601-402M
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1434 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-219-694-13

Query Match 37.3%; Score 86.2; DB 14; Length 1434;
Best Local Similarity 73.7%; Pred. No. 3.1e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;
QY 55 GGAAGCAGCTCTGATTGTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGGAGAAAAC 114
DB 900 GGGGACATGTCTATCTTCATGAAGTTTCAGTCTCAGTCTCCCTTCCCGGAGAAAAC 959
QY 115 GACACGGGAGCTGCGCCAAAGCTCTCTGGTGA-----TGAGCCTAAGGGATGTTT 167
DB 960 GACACGGGAGCAGTCAAGGTTCTCTGGTAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019
QY 168 GTGATGGGCCCTTATCTTGCACACATGGGATCAGACCTCTACCTTCAACC 218
DB 1020 GTACATGGCTCTTACCTACACATGGGATTTGACCTCTATCTCCACTC 1070

RESULT 10

US-09-799-462A-14
; Sequence 14, Application US/09799462A
; Patent No. US20020160970A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,462A
; FILING DATE: 10-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.


```
;
;
;   REGISTRATION NUMBER: 33,779
;   REFERENCE/DOCKET NUMBER: 24601-402G
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858-450-8403
;   TELEFAX: 858-587-5360
;   TELEX: <Unknown>
;
;   INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1400 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: <Unknown>
;   ORIGINAL SOURCE:
;   SEQUENCE DESCRIPTION: SEQ ID NO: 14:
;
; US-09-799-462A-14
;
; Query Match          35.2%   Score 81.4; DB 10; Length 1400;
; Best Local Similarity 71.9%; Pred. No. 1.7e-19;
; Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
;
; QY      55 GGAAGGCACCTCTGATTGCATGAAGGTTCAAGTTCAGTTCCTAGTTCCTCCCGGAGGAAAAAC 114
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAGTGTCTTCTCCCGGAGGAAAAAC 830
;
; QY      115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAAGGGATGGTTT 167
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      831 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCCTAAAGGCTAATCCT 890
;
; QY      168 GTGATGGCCCTTATGCTTGCACACTGGGATGAGACCTCTACCTTACCTTCCACC 218
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      891 GTACATGGCTCTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941
;
;
; RESULT 11
; US-09-836-911A-14
;   Sequence 14, Application US/09836911A
;   Publication No. US20030033617A1
;   GENERAL INFORMATION:
;   APPLICANT: Hadlaczky, Gyula
;           Szalay, Aladar
;   TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
;           AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
;   NUMBER OF SEQUENCES: 34
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Heller Ehrman White & McAuliffe
;   STREET: 4350 La Jolla Village Drive, 6th Floor
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/836,911A
;   FILING DATE: 17-Apr-2002
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/835,682
;   FILING DATE: 10-APR-1997
;   APPLICATION NUMBER: 08/695,191
;   FILING DATE: 07-AUG-1996
;   APPLICATION NUMBER: 08/682,080
;   FILING DATE: 15-JUL-1996
;   APPLICATION NUMBER: 08/629,822
;   FILING DATE: 10-APR-1996
;   ATTORNEY/AGENT INFORMATION:
```

```
;
;
;   NAME: Seidman, Stephanie L
;   REGISTRATION NUMBER: 33,779
;   REFERENCE/DOCKET NUMBER: 24601-402I
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 858-450-8403
;   TELEFAX: 858-587-5360
;   TELEX: <Unknown>
;
;   INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1400 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: Genomic DNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: <Unknown>
;   ORIGINAL SOURCE:
;   SEQUENCE DESCRIPTION: SEQ ID NO: 14:
;
; US-09-836-911A-14
;
; Query Match          35.2%   Score 81.4; DB 11; Length 1400;
; Best Local Similarity 71.9%; Pred. No. 1.7e-19;
; Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
;
; QY      55 GGAAGGCACCTCTGATTGCATGAAGGTTCAAGTTCAGTTCCTAGTTCCTCCCGGAGGAAAAAC 114
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAGTGTCTTCTCCCGGAGGAAAAAC 830
;
; QY      115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAAGGGATGGTTT 167
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      831 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTGTGAGCCCTAAAGGCTAATCCT 890
;
; QY      168 GTGATGGCCCTTATGCTTGCACACTGGGATGAGACCTCTACCTTACCTTCCACC 218
;         || || || || || || || || || || || || || || || || || || || || || || ||
; Db      891 GTACATGGCTCTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941
;
;
; RESULT 12
; US-10-125-767-14
;   Sequence 14, Application US/10125767
;   Publication No. US20020160410A1
;   GENERAL INFORMATION:
;   APPLICANT: Hadlaczky, Gyula
;   TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
;           METHODS
;           FOR PREPARING ARTIFICIAL CHROMOSOMES
;   NUMBER OF SEQUENCES: 34
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Heller Ehrman White & McAuliffe LLP
;   STREET: 4350 La Jolla Village Drive, 7th Floor
;   CITY: San Diego
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92122
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSeq Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/125,767
;   FILING DATE: 17-Apr-2002
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/724,693
;   FILING DATE: 28-NOV-2000
;   APPLICATION NUMBER: 08/835,682
;   FILING DATE: 10-APR-1997
;   APPLICATION NUMBER: 08/695,191
;   FILING DATE: 07-AUG-1996
;   APPLICATION NUMBER: 08/682,080
;   FILING DATE: 15-JUL-1996
;   APPLICATION NUMBER: 08/629,822
```

```

/ APPLICATION NUMBER: 08/695,191
/ FILING DATE: 07-AUG-1996
/ APPLICATION NUMBER: 08/682,080
/ FILING DATE: 15-JUL-1996
/ APPLICATION NUMBER: 08/629,822
/ FILING DATE: 10-APR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Seidman, Stephanie L
/ REGISTRATION NUMBER: 33,779
/ REFERENCE/DOCKET NUMBER: 24601-402L
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 858-450-8403
/ TELEFAX: 858-587-5360
/ TELEX: <Unknown>
/ INFORMATION FOR SEQ ID NO: 14:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1400 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: Genomic DNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: <Unknown>
/ ORIGINAL SOURCE:
/ SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-151-081-14

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best local similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;

QY 55 GGAAGGCACGCTCGATTGTCATGAAGTTCACTGCTCAGTGTCTCCTTCCCTCCCGCAGGAAAAAC 114
DB 771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAGTGTCTTCTCCTCCAGCAAAC 830

QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGGATGGTTTT 167
DB 831 GACACGGGAGCAGGTGAGGTTGCTCTGGGTAAAGCCTCTGAGCCTAAGAGCTAATCCT 890

QY 168 GTCTAGGGCCCTATGCTTCCACACGTGGGATCAGACCTTACCTTCAACCC 218
DB 891 GTACATGGCTCTTTACCTACACACGTGGGATTTGACCTCTATCTCCACTC 941

RESULT 14
US-10-287-313-14
/ Sequence 14, Application US/10287313
/ Publication No. US20030101480A1
/ GENERAL INFORMATION:
/ APPLICANT: Radlaczky, Gyula
/ TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
/ AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
/ NUMBER OF SEQUENCES: 34
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Heller Ehrman White & McAuliffe
/ STREET: 4250 Executive Square, 7th Floor
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 1.5
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/287,313
/ FILING DATE: 01-No. US20030101480A1-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/724,726

```

;; FILING DATE: 28-NOV-2000
;; APPLICATION NUMBER: 08/835,682
;; FILING DATE: 10-APR-1997
;; APPLICATION NUMBER: 08/695,191
;; FILING DATE: 07-AUG-1996
;; APPLICATION NUMBER: 08/682,080
;; FILING DATE: 15-JUL-1996
;; APPLICATION NUMBER: 08/629,822
;; FILING DATE: 10-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 6669-402N
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 858-450-8403
;; TELEFAX: 858-587-5360
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1400 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: <Unknown>
;; ORIGINAL SOURCE:
;; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-287-313-14

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best Local Similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
QY 55 GGAAGCAGCTGTGATTGATGAGGTTTCAGTGTCTTCCCTTCCCGGAGGAAAAC 114
Db 771 GGGAGACATGTCATCTTCAAGAAGGTTGAGTGTCCAGTGTCTTCCCTCCAGGCAAAAC 830
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 831 GACACGGGAGCAGGTGAGGTTGCTTGGTAAAGCCTGTGAGCCTAAGAGCTAATCCT 890
QY 168 GTGATGGGCCCCATCTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
Db 891 GTACATGGCTCTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 941

RESULT 15

US-10-219-694-14
; Sequence 14, Application US/10219694
; Publication No. US20030108914A1
; GENERAL INFORMATION:
; APPLICANT: Hadlaczkv, Gyula
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP
; STREET: 4350 La Jolla Village Drive, 7th Floor
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/219,694
; FILING DATE: 14-Aug-2002
; CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: <Unknown>
;; FILING DATE: 28-NOV-2000
;; APPLICATION NUMBER: 08/835,682
;; FILING DATE: 10-APR-1997
;; APPLICATION NUMBER: 08/695,191
;; FILING DATE: 07-AUG-1996
;; APPLICATION NUMBER: 08/682,080
;; FILING DATE: 15-JUL-1996
;; APPLICATION NUMBER: 08/629,822
;; FILING DATE: 10-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seidman, Stephanie L
;; REGISTRATION NUMBER: 33,779
;; REFERENCE/DOCKET NUMBER: 24601-402M
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 858-450-8403
;; TELEFAX: 858-587-5360
;; TELEX: <Unknown>
;; INFORMATION FOR SEQ ID NO: 14:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1400 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Genomic DNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: <Unknown>
;; ORIGINAL SOURCE:
;; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-219-694-14

Query Match 35.2%; Score 81.4; DB 14; Length 1400;
Best Local Similarity 71.9%; Pred. No. 1.7e-19;
Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
QY 55 GGAAGCAGCTGTGATTGATGAGGTTTCAGTGTCTTCCCTTCCCGGAGGAAAAC 114
Db 771 GGGAGACATGTCATCTTCAAGAAGGTTGAGTGTCCAGTGTCTTCCCTCCAGGCAAAAC 830
QY 115 GACACGGGAGCTGGCCAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGTTT 167
Db 831 GACACGGGAGCAGGTGAGGTTGCTTGGTAAAGCCTGTGAGCCTAAGAGCTAATCCT 890
QY 168 GTGATGGGCCCCATCTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
Db 891 GTACATGGCTCTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 941

Search completed: October 9, 2003, 23:40:57
Job time : 208 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:06:28 ; Search time 73 Seconds
(without alignments)
1396.706 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 ggggggtcgaggctaagca.....ttcaccatgaggttgctt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86.2	37.3	1434	3	US-08-695-191-13
2	86.2	37.3	1434	3	US-08-682-080-13
3	81.4	35.2	1400	3	US-08-695-191-14
4	81.4	35.2	1400	3	US-08-682-080-14
5	35	15.2	50	4	US-08-585-593A-4
6	29.2	12.6	540	4	US-09-250-609-95
7	29.2	12.6	540	4	US-09-250-611-95
8	28.6	12.4	2692	1	US-07-932-454A-2
9	28	12.1	866	4	US-09-620-312D-682
10	27.6	11.9	1865	3	US-08-766-354A-29
11	27.2	11.8	2297	2	US-08-588-983-21
12	27.2	11.8	2297	2	US-08-588-976-21
13	27	11.7	1054	6	5189147-4
14	26.8	11.6	503	4	US-09-250-609-85
15	26.8	11.6	503	4	US-09-250-611-85
16	26.8	11.6	3213	2	US-08-633-770A-4
17	26.4	11.4	13865	3	US-09-009-217-11
18	26.4	11.4	13865	3	US-09-009-656-11
19	26.2	11.3	71	1	US-08-434-001-210
20	26.2	11.3	71	1	US-08-433-585-210
21	26.2	11.3	71	1	US-08-434-425-210
22	26.2	11.3	71	2	US-08-437-667-210
23	26.2	11.3	71	3	US-08-906-955-210
24	26.2	11.3	71	3	US-08-945-909-210
25	26.2	11.3	71	4	US-09-396-002A-210
26	26.2	11.3	71	5	PCT-US96-06060-210
27	26	11.3	66804	4	US-09-740-041-3

Sequence 1, Appli
Sequence 180, App
Sequence 55, Appl
Sequence 792, App
Sequence 8, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 113, App
Sequence 188, App
Sequence 17, Appl
Sequence 22, Appl
Sequence 1, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 1, Appli

28 26 11.3 536165 4 US-09-214-808-1
29 25.8 11.2 444 3 US-08-928-213B-180
30 25.8 11.2 866 4 US-09-495-050A-55
31 25.8 11.2 3883 4 US-09-620-312D-792
32 25.8 11.2 5865 3 US-09-011-745-8
33 25.8 11.2 9058 3 US-08-913-014A-9
34 25.8 11.2 1830121 4 US-09-557-884-1
35 25.8 11.2 1830121 4 US-09-643-990A-1
36 25.6 11.1 419 3 US-09-214-095D-113
37 25.6 11.1 504 4 US-09-328-475C-188
38 25.6 11.1 4168 4 US-09-266-225D-17
39 25.6 11.1 4279 3 US-09-041-886-22
40 25.6 11.1 319608 4 US-09-539-333D-1
41 25.6 11.1 319608 4 US-09-679-409-1
42 25.4 11.0 4643 4 US-09-436-874-1
43 25.4 11.0 4403765 3 US-09-103-840A-2
44 25.4 11.0 4411529 3 US-09-103-840A-1
45 25.2 10.9 371 1 US-08-620-467A-1

ALIGNMENTS

RESULT 1
US-08-695-191-13
Sequence 13, Application US/08695191
Patent No. 6025155
GENERAL INFORMATION:
APPLICANT: Hadlaczky, Gyula
APPLICANT: Szalay, Aladar
TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/695,191
FILING DATE: 07-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE: 15-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/629,822
FILING DATE: 10-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6869-402C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO

```

: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
US-08-695-191-13

Query Match 37.3% Score 86.2; DB 3; Length 1434;
Best Local Similarity 73.7%; Pred. No. 1.3e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

QY 55 GGAAGGCACGCTCGATTGCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAAAAC 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 900 GGGGACATGTCATCTTTTCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAAAAC 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 115 GACACGGGAGCTGGCCAAAGACCTCTCTGGGTGA-----TGAGCCTTAAGGGATGGTTTT 167
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 960 GACACGGGAGCAGGTCAGGGTTGCTCTGGGTAAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 168 GTGTAGGCGCCCTATGCTTCCACACTGGGATCAGACCTCTACTCTTCACCC 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1020 GTACATGGCTCCCTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 1070
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
US-08-682-080-13
: Sequence 13, Application US/08682080
: Patent No. 6077697
: GENERAL INFORMATION:
: APPLICANT: Hadlaczky, Gyula
: APPLICANT: Szalay, Aladar
: TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND METHODS FOR
: TITLE OF INVENTION: PREPARING ARTIFICIAL CHROMOSOMES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,080
: FILING DATE: 10-APR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/629,822
: FILING DATE: 10-APR-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie L
: REGISTRATION NUMBER: 33,779
: REFERENCE/DOCKET NUMBER: 6869-402B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-238-0999
: TELEFAX: 619-238-0062
: TELEX:
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1434 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
US-08-682-080-13

```

```

Query Match          37.3%; Score 86.2; DB 3; Length 1434;
Best Local Similarity 73.7%; Pred. No. 1.3e-21;
Matches 126; Conservative 0; Mismatches 38; Indels 7; Gaps 1;

QY      55  GGAAGGCACCTCTGATTGCATGAAGGTTCAAGTTCCTAGTTCCCTTCCCGCCAGGAAAAC 114
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      900  GGGGGACATGTCTATCTTTTCATGAAGGTTCAAGTTCCTAGTTCCCTTCCCGCCAGGAAAAC 959
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      115  GACACGGGAGCTGGCCGAAGACCTCTCTCGGCTGA-----TGAGCCCTAAGGGATGGTTTT 167
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      960  GACACGGGAGCAGGTCAGGCTTGCTCTGCGTAAAAAGCCTGTGAGCCTGGGAGCTAATCCT 1019
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      168  GTGTAGGGCCCCCTATGCTTCACACACTGGGGATCAGACCTCTACCTTCACCC 218
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1020  GTACATGGCTCTTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 1070
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-695-191-14
; Sequence 14, Application US/08695191
; Patent No. 6025155
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; APPLICANT: Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF AND
; TITLE OF INVENTION: METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92101-2926
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/695,191
; FILING DATE: 07-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE: 15-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6869-402C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-238-0999
; TELEFAX: 619-238-0062
; TELEX:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-695-191-14

```

Query Match 35.2%; Score 81.4; DB 3; Length 1400;

771	GGGAGACATGTCATCTTTTCAAGAAGTTGAGTGTCTCAAGTGTCTCTCTCCAGGCCAAAC	830
Db		
115	GACACGGAGCTGGCCAAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTTT	167
QY		
831	GACACGGGAGAGATGTAGGGTTGCTCTGGGTAAAGCCTGTGAGCCTTAAGAGCTAATCCT	890
Db		
168	GTGTAGGGCCCTATGCTTGACACTGGGGATCAGACCTCTACCTTCACCC	218
QY		
891	GTACATGGCTCTTTTACCTACACACTGGGATTTGACCTTATCTGCACATC	941
Db		

RESULT 5
US-08-581
; Sequen
; Patent

```

/ GENERAL INFORMATION:
/ APPLICANT:  ABKEN, Hinrich J
/ APPLICANT:  ALBERT, Winfried
/ APPLICANT:  JUNGFER, Herbert
/ TITLE OF INVENTION:  METHOD OF IDENTIFYING HUMAN AND ANIMAL
/ TITLE OF INVENTION:  CELLS CAPABLE OF UNLIMITED PROLIFERATION OR TUMOR
/ TITLE OF INVENTION:  FORMATION
/ NUMBER OF SEQUENCES:  66
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSES:  Nikaido, Marmelstein, Murray & Oram LLP
/ STREET:  655 Fifteenth Street N.W. Suite 330
/ CITY:  Washington
/ STATE:  D.C.
/ COUNTRY:  U.S.A.
/ ZIP:  20005-5701

```

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/585,593A
 FILING DATE: 16-JAN-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/EP94/02307
 FILING DATE: 13-JUL-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: DE P 43 23 727.4
 FILING DATE: 15-JUL-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Murray, Robert B.
 REGISTRATION NUMBER: 22,980
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)638-5000
 TELEFAX: (202)638-4810
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 50 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA

1 GATCAGACCTCTACCTTCACCCATGAGGGCTTGCTT 35

```

; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210002
; CURRENT APPLICATION NUMBER: US/09/250,609A
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-609-95

Query Match      12.6%; Score 29.2; DB 4; Length 540;
Best Local Similarity 54.7%; Pred. No. 0.54;
Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 26 CAGAGGATAGCTGCTGTTGGCATCCTGTGGAAGGACGCTGCTGATTCGATGAAGGTTGAG 85
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 CAGAACCTGTCTAGGAGCTGGCATGATGTGCAGGTCTCTACTGCTTACAGAGAGACTCAA 286
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 86 TGTCTTAGTTCCCTTCCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 131
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 GAAACTCTTTCACAGGCTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-09-250-611-95
; Sequence 95, Application US/09250611
; Patent No. 6528283
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; APPLICANT: Bassett, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 95
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-611-95

Query Match      12.6%; Score 29.2; DB 4; Length 540;
Best Local Similarity 54.7%; Pred. No. 0.54;
Matches 58; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 26 CAGAGGATAGCTGCTGTTGGCATCCTGTGGAAGGACGCTGCTGATTCGATGAAGGTTGAG 85
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 CAGAACCTGTCTAGGAGCTGGCATGATGTGCAGGTCTCTACTGCTTACAGAGAGACTCAA 286
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 86 TGTCTTAGTTCCCTTCCCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 131
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 GAAACTCTTTCACAGGCTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 332
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-07-932-454A-2/c
; Sequence 2, Application US/07932454A
; Patent No. 5262318
; GENERAL INFORMATION:
; APPLICANT: GUTHRIE, ELLEN P.
; TITLE OF INVENTION: ISOLATED DNA ENCODING THE Sphi
; TITLE OF INVENTION: RESTRICTION ENDONUCLEASE AND RELATED METHODS FOR PRODUCING
; TITLE OF INVENTION: THE SAME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET

```

```

; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/932,454A
; FILING DATE: 19920820
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: 42078
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2692 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 703..1653
; OTHER INFORMATION: /note= "METHYLASE GENE STARTS AT
; OTHER INFORMATION: POSITION 703/ENDS AT 1653. RESTRICTION
; OTHER INFORMATION: ENDONUCLEASE STARTS AT POSITION 1703/ENDS AT 2410"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1703..2410
; US-07-932-454A-2

Query Match      12.4%; Score 28.6; DB 1; Length 2692;
Best Local Similarity 55.6%; Pred. No. 1.8;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 53 GTGGAAGGCACGCTGCTGATTCGATGAAGGTTGATGCTGCTGCTGCTGCTGCTGCTGCTG 151
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 418 GGGGGAGTTTCCACCGATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 320
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 ACGACACGGGAGCTGGCCAGACGCTCTCTGGGTGATGAG 151
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 358 GCGGTACGAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 320
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-620-312D-682/c
; Sequence 682, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.

```


Db 2196 CACTGAGGCTGCTTCTCCCTGGGTCAGGGACCTAGTGTGGAGATGATTCCTGCT 2137
QY 143 GGTGATGAGCTAAGGATGTTTTGTGTAGGCCCCCTAT---GCTTGACACATGGGGAT 199
Db 2136 TGAATAGTCAGATGAGGCTCCCTTCTGTGTCCTTAGACAGGAGCTCTGGGCT 2077
QY 200 CAGACCTCTACCTTCACCCATGAGGCTT 227
Db 2076 CTAACCTCTGCACTCTCTCTGGGAGGTT 2049

RESULT 12

US-08-588-976-21/c
; Sequence 21, Application US/08588976
; Patent No. 5891717
; GENERAL INFORMATION:
; APPLICANT: Christopher B. Newgard, et al.
; TITLE OF INVENTION: Methods and Compositions for
; TITLE OF INVENTION: Inhibiting Hexokinase
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: US
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/588,976
; FILING DATE: Concurrently herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: UTSD:481/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-588-976-21

Query Match 11.8%; Score 27.2; DB 2; Length 2297;
Best Local Similarity 49.0%; Pred. No. 5.4; Mismatches 103; Indels 3; Gaps 1;
Matches 102; Conservative 0;
QY 23 GCACAGAGATGAGCTTGTGTGTCATCTCTGGAAGGCACGCTGATGTCATGAAGTT 82
Db 2256 GTACAGAGCTGGCATTTGGGAACGAGGAGAGAGGACCTGGTGTCTTGGAGGA 2197
QY 83 CAGTGTCTAGTTCCTTCCCTCCAGGAAACGACAGGGAGCTGGCCACAGACCTCTCTG 142
Db 2196 CACTGAGGGTGGCTTCTTCTCTCTGCTGAGGACCTGAGTGTGGAGATGATTCCTGCT 2137
QY 143 GGTGATGAGCTAAGGATGTTTTGTGTAGGCCCCCTAT---GCTTGACACATGGGAT 199
Db 2136 TGAATAGTCAGATGAGGCTCCCTTCTGTGTCCTTAGACAGGAGCTCTGGGCT 2077
QY 200 CAGACCTCTACCTTCACCCATGAGGCTT 227
Db 2076 CTAACCTCTGCACTCTCTCTGGGAGGTT 2049

RESULT 13
5189147-4
; Patent No. 5189147
; APPLICANT: SAITO, HARUO; KRANZ, DAVID M.; EISEN, HERMAN N.;
; TONEGAWA, SUSUMU
; TITLE OF INVENTION: METEORODIMERIC T LYMPHOCYTE RECEPTOR
; ANTIBODY
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/271,216
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 666,988
; FILING DATE: 31-OCT-1984
; APPLICATION NUMBER: 620,122
; FILING DATE: 13-JUN-1984
; SEQ ID NO: 4
; LENGTH: 1054
5189147-4

Query Match 11.7%; Score 27; DB 6; Length 1054;
Best Local Similarity 53.3%; Pred. No. 4.5;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 47 CATCCTGTGGAAGGCACGCTGTGATTCATGAAGGTTGAGTTCAGTTCCTCTAGTTCCTTCCCCCA 106
Db 803 CCTACTGGGAAGGCCACCTATATCTGTGCTGTCAGTGGCTAGTGTGATGGCCAT 862
QY 107 GGAAGAACGACACGGGAGCTGGCCAGCAAGACCTCTCTGGGTGATGAGCC 153
Db 863 GGTCAAGAAAAAATTCCTGAGACAAAATTTATGTCATCCTGAGCC 909

RESULT 14

US-09-250-609-85
; Sequence 85, Application US/09250609A
; Patent No. 6458943
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210002
; CURRENT APPLICATION NUMBER: US/09/250,609A
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-250-609-85

Query Match 11.6%; Score 26.8; DB 4; Length 503;
Best Local Similarity 53.9%; Pred. No. 3.9;
Matches 55; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 26 CAGAGATGATGCTGCTGTGGCATCTCTGGAAGGCACGCTGATGTCATGAAGTTTCA 85
Db 376 CAGAACCTGTCTAGGACCTGGCATGATGTCAGTCTCTACTGCTCTACAGNAGACTCA 435
QY 86 TGTCTAGTTCCTTCCCTCCAGGAAAAACACACGGGAGCTG 127
Db 436 GAAACTCTTTCACAGGCTGGACAGAAAACATCAGCTGCCCTG 477

RESULT 15

US-09-250-611-85
; Sequence 85, Application US/09250611
; Patent No. 6528283
; GENERAL INFORMATION:
; APPLICANT: Byrne, Jennifer A.
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family

```

; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-250-611-85

```

```

Query Match      11.6%; Score 26.8; DB 4; Length 503;
Best Local Similarity 53.9%; Pred No. 3.9;
Matches 55; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY      26 CAGAGGATAGTTGCTGTGGCATCCTGTGGAGGCACGCTGTGATTGCATGAAGGTTTCAG 85
      |||| | || | ||||| |||| | || | || |||| |||
Db      376 CAGAACCTGCTTAGGAGCTGGCATGATGTGCAGGTTCTACTGCCTACAAGAAGACTCAA 435

QY      86 TGTCCTAGTTCCCTTCCCCCAGGAAAAAGCACACGGGAGCTG 127
      || |||| | || | |||| || |||
Db      436 GAACTCTTTTCACAGGCTGGACAGAAAAACATCAGCTGCCCTG 477

```

Search completed: October 9, 2003, 22:44:00
Job time : 81 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:19:09 ; Search time 3003 Seconds
(without alignments)
2551.060 Million cell updates/sec

Title: US-09-673-716-1
Perfect score: 231
Sequence: 1 gtaggggtcgaggctaagca.....ttcaccatgaggttgctt 231

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 33363688 seqs, 16581889874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents_NA_Main:*

- 1: /cgn2_6/ptodata/1/pna/PTCTUS_COMB.seq:*
- 2: /cgn2_6/ptodata/1/pna/PTCTUS_COMB.seq.oid:*
- 3: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
- 4: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
- 5: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
- 6: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
- 7: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
- 8: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
- 9: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
- 10: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
- 11: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
- 12: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
- 13: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
- 14: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
- 15: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
- 16: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
- 17: /cgn2_6/ptodata/1/pna/US092A_COMB.seq:*
- 18: /cgn2_6/ptodata/1/pna/US092B_COMB.seq:*
- 19: /cgn2_6/ptodata/1/pna/US093A_COMB.seq:*
- 20: /cgn2_6/ptodata/1/pna/US093B_COMB.seq:*
- 21: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
- 22: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
- 23: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
- 24: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
- 25: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
- 26: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
- 27: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
- 28: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
- 29: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
- 30: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
- 31: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
- 32: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
- 33: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
- 34: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
- 35: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
- 36: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
- 37: /cgn2_6/ptodata/1/pna/US098D_COMB.seq:*
- 38: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
- 39: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
- 40: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
- 41: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
- 42: /cgn2_6/ptodata/1/pna/US099E_COMB.seq:*
- 43: /cgn2_6/ptodata/1/pna/US099F_COMB.seq:*

44: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*

45: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*

46: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*

47: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*

48: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*

49: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*

50: /cgn2_6/ptodata/1/pna/US103A_COMB.seq:*

51: /cgn2_6/ptodata/1/pna/US103B_COMB.seq:*

52: /cgn2_6/ptodata/1/pna/US104A_COMB.seq:*

53: /cgn2_6/ptodata/1/pna/US104B_COMB.seq:*

54: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*

55: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*

56: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*

57: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*

58: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*

59: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*

60: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*

61: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*

62: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*

63: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*

64: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*

65: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*

66: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*

67: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*

68: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*

69: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*

70: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*

71: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*

72: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*

73: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*

74: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*

75: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*

76: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*

77: /cgn2_6/ptodata/1/pna/US6023A_COMB.seq:*

78: /cgn2_6/ptodata/1/pna/US6023B_COMB.seq:*

79: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*

80: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*

81: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*

82: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*

83: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*

84: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*

85: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*

86: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*

87: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*

88: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*

89: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*

90: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*

91: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*

92: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*

93: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*

94: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*

95: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*

96: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*

97: /cgn2_6/ptodata/1/pna/US6042_COMB.seq:*

98: /cgn2_6/ptodata/1/pna/US6043_COMB.seq:*

99: /cgn2_6/ptodata/1/pna/US6044_COMB.seq:*

100: /cgn2_6/ptodata/1/pna/US6045_COMB.seq:*

101: /cgn2_6/ptodata/1/pna/US6046_COMB.seq:*

102: /cgn2_6/ptodata/1/pna/US6047_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	231	100.0	231	29	US-09-673-716-1 Sequence 1, Appl1
2	231	100.0	391	29	US-09-673-716-4 Sequence 4, Appl1
3	231	100.0	2052	46	US-10-144-771-25944 Sequence 25944, A
4	231	100.0	2052	91	US-60-360-207-25944 Sequence 25944, A


```
RESULT 3
US-10-144-771-25944
; Sequence 25944, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25944
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-25944

Query Match      100.0%; Score 231; DB 46; Length 2052;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1088 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1147

QY 61 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1148 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1207

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 180
Db 1208 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 1267

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1318

RESULT 4
US-60-360-207-25944
; Sequence 25944, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25944
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-25944

Query Match      100.0%; Score 231; DB 91; Length 2052;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1088 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1147

QY 61 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1148 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1207

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 180
Db 1208 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 1267

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1318
```

```
Db 1268 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1318

RESULT 5
US-10-144-771-25350
; Sequence 25350, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25350
; LENGTH: 2098
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-25350

Query Match      100.0%; Score 231; DB 46; Length 2098;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1201 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1260

QY 61 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1261 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1320

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 180
Db 1321 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 1380

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
Db 1381 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1431

RESULT 6
US-60-360-207-25350
; Sequence 25350, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25350
; LENGTH: 2098
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-25350

Query Match      100.0%; Score 231; DB 91; Length 2098;
Best Local Similarity 100.0%; Pred. No. 6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 60
Db 1201 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTTCTGTGGCATCCTGTGGAAGG 1260

QY 61 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
Db 1261 CACGCTCTGATTGCATGAAGGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 1320

QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 180
Db 1321 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGGCTTAAGGATGTTTGTAGGGCCCT 1380

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
```


|||||
Db 1381 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1431

RESULT 7
US-10-144-771-33482
; Sequence 33482, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 33482
; LENGTH: 2258
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-33482

Query Match 100.0%; Score 231; DB 46; Length 2258;
Best Local Similarity 100.0%; Pred. No. 6.2e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 60
Db 1299 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 1358

QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 120
Db 1359 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 1418

QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 180
Db 1419 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 1478

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1479 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1529

RESULT 8
US-60-360-207-33482
; Sequence 33482, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 33482
; LENGTH: 2258
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-33482

Query Match 100.0%; Score 231; DB 91; Length 2258;
Best Local Similarity 100.0%; Pred. No. 6.2e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 60
Db 1299 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 1358

QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 120
Db 1359 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 1418

QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 180
Db 1419 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 1478

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1479 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1529

RESULT 9
US-10-144-771-22182
; Sequence 22182, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 22182
; LENGTH: 2353
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-22182

Query Match 100.0%; Score 231; DB 46; Length 2353;
Best Local Similarity 100.0%; Pred. No. 6.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 60
Db 1328 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 1387

QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 120
Db 1388 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 1447

QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 180
Db 1448 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 1507

QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1508 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1558

RESULT 10
US-60-360-207-22182
; Sequence 22182, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 22182
; LENGTH: 2353
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-22182

Query Match 100.0%; Score 231; DB 91; Length 2353;
Best Local Similarity 100.0%; Pred. No. 6.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 60
Db 1328 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTCTGGCATCTCTGTGGAAGG 1387

QY 61 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 120
Db 1388 CACGCTCTGATTGATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACG 1447

QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 180
Db 1448 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGGTTTGTGTAGGGCCCT 1507

QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
|||||
Db 1508 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1558
|||||
RESULT 11
US-10-144-771-25153
; Sequence 25153, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25153
; LENGTH: 3403
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-25153
Query Match 100.0%; Score 231; DB 46; Length 3403;
Best Local Similarity 100.0%; Pred. No. 7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 60
|||||
Db 1537 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 1596
|||||
QY 61 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 120
|||||
Db 1597 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 1656
|||||
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
|||||
Db 1657 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1716
|||||
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
|||||
Db 1717 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1767
|||||

RESULT 12
US-60-360-207-25153
; Sequence 25153, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 25153
; LENGTH: 3403
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-25153
Query Match 100.0%; Score 231; DB 91; Length 3403;
Best Local Similarity 100.0%; Pred. No. 7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 60
|||||
Db 1537 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 1596
|||||
QY 61 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 120
|||||
Db 1597 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 1656
|||||
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
|||||

Db 1657 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1716
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
|||||
Db 1717 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1767
|||||
RESULT 13
US-10-144-771-28382
; Sequence 28382, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 28382
; LENGTH: 5633
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-28382
Query Match 100.0%; Score 231; DB 46; Length 5633;
Best Local Similarity 100.0%; Pred. No. 8.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 60
|||||
Db 1641 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 1700
|||||
QY 61 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 120
|||||
Db 1701 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 1760
|||||
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
|||||
Db 1761 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 1820
|||||
QY 181 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 231
|||||
Db 1821 ATGCTTGCACACTGGGATCAGACCTCTACCTTACCCATGAGGCTTGCTT 1871
|||||

RESULT 14
US-60-360-207-28382
; Sequence 28382, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 28382
; LENGTH: 5633
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-28382
Query Match 100.0%; Score 231; DB 91; Length 5633;
Best Local Similarity 100.0%; Pred. No. 8.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 60
|||||
Db 1641 GTGGGGTGGAGGCTAAGCAGCTGCACAGGATAGCTTCTGTGGCATCTCTGTGGAAGG 1700
|||||
QY 61 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 120
|||||
Db 1701 CACGCTGATTGCATGAAGGTTTCAGTGTCTAGTCTCCCTCCCGAGGAAACGACAG 1760
|||||
QY 121 GGAGCTGGCCAAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
|||||

Db 1761 GGAGCTGCCAAGACCTCTCTGGGTGATGACCTTAAGGATGGTTTGTGTAGGGCCCT 1820
Qy 181 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1821 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1871

RESULT 15

US-10-144-771-28138
; Sequence 28138, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 28138
; LENGTH: 5675
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(5675)
; OTHER INFORMATION: n = A,T,C or G
US-10-144-771-28138

Query Match 100.0%; Score 231; DB 46; Length 5675;
Best Local Similarity 100.0%; Pred. No. 8.3e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGGGGTGCAGGGCTAAGCACTGCACAGAGATAGCTTGCTTGGCATCCTGTGGAAGG 60
Db 1301 GTGGGGTGCAGGGCTAAGCACTGCACAGAGATAGCTTGCTTGGCATCCTGTGGAAGG 1360
Qy 61 CACGCTCTGATTGCATGAGGTTCAAGTCTCCTAGTTCCTCCCTCCCGAGGAAAAAGCACAG 120
Db 1361 CACGCTCTGATTGCATGAGGTTCAAGTCTCCTAGTTCCTCCCTCCCGAGGAAAAAGCACAG 1420
Qy 121 GGAGCTGGCCAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGGTTTGTGTAGGGCCCT 180
Db 1421 GGAGCTGGCCAAGACCTCTCTGGGTGATGAGCCTTAAGGGATGGTTTGTGTAGGGCCCT 1480
Qy 181 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 1481 ATGCTTGACACTGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 1531

Search completed: October 9, 2003, 23:34:11
Job time : 3004 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 22:41:03 ; Search time 171 Seconds
(without alignments)
1886.736 Million cell updates/sec

Title: US-09-673-716-1
Perfect score: 231
Sequence: 1 gtgggggagagctaaagca.....ttcaccatgaggtgctt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1776535 seqs, 698337343 residues

Total number of hits satisfying chosen parameters: 3553070

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_New.*

- 1: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
- 7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	72	31.2	96593	1	PCT-US02-38582-151
2	35	15.2	30593	7	US-60-495-114-16231
3	34	14.7	49596	7	US-60-495-114-16231
4	32.6	14.1	1106	6	US-10-425-114A-16270
5	32.6	14.1	1145	6	US-10-425-114A-16270
6	31	13.4	1205	6	US-10-425-114A-16110
7	31	13.4	1211	6	US-10-425-114A-23937
8	30.2	13.1	25995	7	US-60-500-337-19598
9	29.8	12.9	37637	7	US-60-500-315-11368
10	29.6	12.8	959	6	US-10-663-561-452
11	29.6	12.8	1531	6	US-10-425-114A-1045
12	29.6	12.8	2731748	6	US-10-297-465B-1
13	29.4	12.7	19946	7	US-60-500-337-19711
14	28.8	12.5	34227	7	US-60-500-315-11594
15	28.4	12.3	25821	7	US-60-487-610-20089
16	28.4	12.3	25821	7	US-60-485-450-12348
17	28.2	12.2	15448	7	US-60-500-315-11760
18	27.8	12.0	201	7	US-60-485-450-10495
19	27.8	12.0	201	7	US-60-485-450-10505
20	27.8	12.0	201	7	US-60-485-450-10515
21	27.8	12.0	201	7	US-60-485-450-10518
22	27.8	12.0	34570	1	PCT-US02-38582-157
23	27.8	12.0	45945	7	US-60-500-337-19316
24	27.8	12.0	85568	7	US-60-495-114-16789
25	27.8	12.0	85568	7	US-60-495-114-16789
26	27.8	12.0	329776	5	US-09-947-914-49

Sequence 110590,
Sequence 75, Appl
Sequence 10501, A
Sequence 10511, A
Sequence 10521, A
Sequence 10521, A
Sequence 110590,
Sequence 222, App
Sequence 772, App
Sequence 774, App
Sequence 773, App
Sequence 1, Appl
Sequence 221, App
Sequence 3, Appl
Sequence 1, Appl
Sequence 12242, A
Sequence 20005, A
Sequence 220, App
Sequence 5, Appl

ALIGNMENTS

RESULT 1

PCT-US02-38582-151
; Sequence 151, Application PC/TUS0238582
; GENERAL INFORMATION:
; APPLICANT: SAGRES DISCOVERY
; APPLICANT: MORRIS, David W.
; APPLICANT: ENGELHARD, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000143
; CURRENT APPLICATION NUMBER: PCT/US02/38582
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/997,722
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 96593
; TYPE: DNA
; ORGANISM: Mus Musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(96593)
; OTHER INFORMATION: n = A,T,C or G
PCT-US02-38582-151

Query Match 31.2%; Score 72; DB 1; Length 96593;
Best Local Similarity 72.1%; Pred. No. 6.7e-14;
Matches 124; Conservative 0; Mismatches 40; Indels 8; Gaps 2;
QY 55 GGAAGGCACGTCGTGATTGTCATGAGGTCAGTTCCTAGTTCCTTCCTTCCTCCAGGCAAA 113
Db 1044 GCGACATGTCATCTCTTCAAGAAGGTTGAGTCCAGGTCCTCTCTCCAGGCAAA 1103
QY 114 CGACAGGGAGCTGGCCAGACCTCTCTGGGTG-----TGAGCCTAGGATGGTTT 166
Db 1104 CGACAGGGAGGAGGTCAGGTTGCTCTGGGTAAGCCCTGTGAGCCTAAGAGCTAATCC 1163
QY 167 TGTGTAGGGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTCAACC 218
Db 1164 TGTACATGGCTCTCTTACCTGACACTGGGGATTTGACCTCTATCTCCATC 1215

RESULT 2

US-60-495-114-16231
; Sequence 16231, Application US/60495114
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001480
; CURRENT APPLICATION NUMBER: US/60/495,114
; CURRENT FILING DATE: 2003-08-15
; NUMBER OF SEQ ID NOS: 91238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16231
; LENGTH: 30593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-495-114-16231

Query Match 15.2%; Score 35; DB 7; Length 30593;
Best Local Similarity 55.3%; Pred. No. 0.21;
Matches 68; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
QY 4 GGGTGGCAGGTAAGCACTGCACAGAGGATAGCTTGTGTGGCATCTCTGTGGAGGCAC 63
DB 15181 GGGTGGGGACACTCAGAGCCAGAGACTCAGAGCTCTGTGCCATGGCCCAAGCCCT 15240
QY 64 GTCGTATGCATGAAGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGACACGGGA 123
DB 15241 GTTGGATGTTGGCATCTGCCCTGCCAGGACCCCGCCAGCCAGGACAGACAGAG 15300
QY 124 GCT 126
DB 15301 GCT 15303

RESULT 3
US-60-495-135-3644/c
; Sequence 3644, Application US/60495135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN ENZYME PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001479
; CURRENT APPLICATION NUMBER: US/60/495,135
; CURRENT FILING DATE: 2003-08-15
; NUMBER OF SEQ ID NOS: 18339
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3644
; LENGTH: 49596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-495-135-3644

Query Match 14.7%; Score 34; DB 7; Length 49596;
Best Local Similarity 57.5%; Pred. No. 0.52;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 11 AGGCTAAGCACTGCACAGAGATAGCTTGTGTGGCATCTCTGTGGAGGCACGCTCTGAT 70
DB 4553 AGGCAGACTCGGGACACTACTCTGTGTACCGCATATTCAGCTGGGACTCTTAC 4494
QY 71 TGCATGAAGGTTTCAGTCTCCTAGTTCCTTCCCGCAGGAAAAACGA 116
DB 4493 TTCAGCCCTTCTCCTTGTCCAGTTCCTTGGCAGGCAGATCCA 4448

RESULT 4
US-10-425-114A-16270
; Sequence 16270, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 16270
; LENGTH: 1106
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3062-025-A9_FLI
US-10-425-114A-16270

Query Match 14.1%; Score 32.6; DB 6; Length 1106;
Best Local Similarity 52.6%; Pred. No. 0.51;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 31 GATAGCTTGTGTGGCATCTCTGTGGAGGACGCTCTGATTCGATGAAGGTTCAAGTGTC 90
DB 715 GCTAGAATGGTCCAGCCCAACGCGGAAAGCTGGTATGGTTTCTGGACAAGCGCGGCA 774
QY 91 TAGTTCCTTCCCGCAGGAAAAACGACACGGGAGCTGGCCAAAGACCTCTCTGGGTGATGA 150
DB 775 TCATTCCTTGACTCCAGACGACGACGACGACGCTTGTACGATCGTCACAGTGACTA 834
QY 151 GCCTAAGGGATGGTT 165
DB 835 ACTCTGAGGATCGTT 849

RESULT 5
US-10-425-114A-25489
; Sequence 25489, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 25489
; LENGTH: 1145
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3912-015-C11_FLI
US-10-425-114A-25489

Query Match 14.1%; Score 32.6; DB 6; Length 1145;
Best Local Similarity 52.6%; Pred. No. 0.52;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 31 GATAGCTTGTGTGGCATCTCTGTGGAGGACGCTCTGATTCGATGAAGGTTCAAGTGTC 90
DB 753 GCTAGAATGGTCCAGCCCAACGCGGAAAGCTGGTATGGTTTCTGGACAAGCGCGGCA 812
QY 91 TAGTTCCTTCCCGCAGGAAAAACGACACGGGAGCTGGCCAAAGACCTCTCTGGGTGATGA 150
DB 813 TCATTCCTTGACTCCAGACGACGACGACGACGCTTGTACGATCGTCACAGTGACTA 872
QY 151 GCCTAAGGGATGGTT 165
DB 873 ACTCTGAGGATCGTT 887

RESULT 6
US-10-425-114A-16110

; Sequence 16110, Application US/10425114A
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 16110

; LENGTH: 1205

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3061-082-A12_FLI

US-10-425-114A-16110

Query Match 13.4%; Score 31; DB 6; Length 1205;

Best Local Similarity 51.9%; Pred. No. 1.8;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 31 GATAGCTTGCTTGGCATCTCTGTGAAGGCACGCTGATTCATGCAAGGTTTCAGTGTC 90

DB 815 GCTAGAAATGGTCCAGCCACGCGGAAGCTGGTATGTTCTGGACAAGCGCGGCA 874

QY 91 TAGTTCCTTCCCCCAGGAAAACGACACGGGAGCTGGCCAGACCTCTCTGGGTGATGA 150

DB 875 TCATTCTCTTGATCTCAAGAACGACGATGACGACGCTTGCTACGAGCGTCATGAGTGACTA 934

QY 151 GCCTAAGGATGGTT 165

DB 935 ACTCGAGGATCCTT 949

RESULT 7

US-10-425-114A-23937

; Sequence 23937, Application US/10425114A

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 23937

; LENGTH: 1211

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3606-028-C5_FLI

US-10-425-114A-23937

Query Match 13.4%; Score 31; DB 6; Length 1211;

Best Local Similarity 51.9%; Pred. No. 1.8;

Matches 70; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 31 GATAGCTTGCTTGGCATCTCTGTGAAGGCACGCTGATTCATGCAAGGTTTCAGTGTC 90

DB 787 GCTAGAAATGGTCCAGCCACGCGGAAGCTGGTATGTTCTGGACAAGCGCGGCA 846

QY 91 TAGTTCCTTCCCCCAGGAAAACGACACGGGAGCTGGCCAGACCTCTCTGGGTGATGA 150

DB 151 GCTAAGGATGGTT 165

DB 935 ACTCGAGGATCCTT 949

DB 847 TCATTCTCTTGACTCCAGAACGACGATGACGAGCTTGCTACGAGCGTCATGAGTGACTA 906

QY 151 GCCTAAGGATGGTT 165

DB 907 ACTCGAGGATCCTT 921

RESULT 8

US-60-500-337-19598

; Sequence 19598, Application US/60500337

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND

; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: CL001483

; CURRENT APPLICATION NUMBER: US/60/500,337

; CURRENT FILING DATE: 2003-09-05

; NUMBER OF SEQ ID NOS: 123188

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19598

; LENGTH: 25995

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-500-337-19598

Query Match 13.1%; Score 30.2; DB 7; Length 25995;

Best Local Similarity 53.9%; Pred. No. 8.6;

Matches 62; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 59 GGCAGCTGTGATTCATGAAGTTCAGTCTCAGTTCCTTCCTCCCGCCAGAAAAGGACA 118

DB 3083 GGCAGGTGCGAGGGGCTCAGGGTTCACCTCTGCTGGCTCGGCTGACCCCTCTGGCTCTCA 3142

QY 119 CGSGAGCTGCGCAAGACCTCTCTGGTGATGACCTAAGCGATGGTTTGTGTAG 173

DB 3143 CAGGCGCTGGGACTGCGCGGGCTGGTGGGGAGGCGGAGGACTCGGAGGGGGAAG 3197

RESULT 9

US-60-500-315-11368

; Sequence 11368, Application US/60500315

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS

; TITLE OF INVENTION: OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001484

; CURRENT APPLICATION NUMBER: US/60/500,315

; CURRENT FILING DATE: 2003-09-05

; NUMBER OF SEQ ID NOS: 69978

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11368

; LENGTH: 37637

; TYPE: DNA

; ORGANISM: Homo sapiens

US-60-500-315-11368

Query Match 12.9%; Score 29.8; DB 7; Length 37637;

Best Local Similarity 63.0%; Pred. No. 13;

Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 51 CTGTGGAAGGCACGCTGCTGATTCATGAAGGTTTCAGTGTCTAGTTCCTCCCTCCCGCAGGAA 110

DB 36010 CTCITGGGAGGAAGTCTGTTTGAATGTAGGCGTTGTGTATATTTCCCTTCACACAGGA 36069

QY 111 AAACGACACGGGA 123

DB 36070 TCACGCTCACTGTA 36082

RESULT 10

US-10-663-561-452

```

; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medjanis, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; PRIOR FILING DATE: 2001-06-17
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

Query Match 12.8%; Score 29,6; DB 6; Length 2731748;
Best Local Similarity 64.7%; Pred. No. 52;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 68 GATTGCATGAAGGTTTCAGTGCTCCTTCCCTCCAGGAAAAACACACGGGAGCTG 127
Db 719549 GTTTGGATGAACCTCCAGAGTGCTCGTCGCTTCGCGCGGAAAAATCGCGCGCGCGGTT 719608

QY 128 GCCAAGAC 135
Db 719609 CCCGCGTC 719616

RESULT 13
US-60-500-337-19711/c
; Sequence 19711, Application US/60500337
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001483
; CURRENT APPLICATION NUMBER: US/60/500,337
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 123188
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19711
; LENGTH: 19946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-337-19711

Query Match 12.7%; Score 29,4; DB 7; Length 19946;
Best Local Similarity 54.1%; Pred. No. 15;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 67 TGATTGCATGAAGGTTTCAGTGCTCCTTCCCTCCAGGAAAAACACACGGGAGCT 126
Db 662 TCATGCGGGAAGGGCGGCGCTCGAGGCTCCCTCCCGCCGCCCATCATCTGTGGGCT 603

QY 127 GGCCACAGACCTCTCGGTGTATGAGCCTAAGGGATGTTTGTGTAGGGCC 177
Db 502 GGCCCCAGCATAGCTGGGAGGAGGAGCTGTGCTTGCTGAGCCCTCGTGC 552

RESULT 14
US-60-500-315-11594
; Sequence 11594, Application US/60500315
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN G-PROTEIN COUPLED RECEPTOR PROTEINS, METHODS
; TITLE OF INVENTION: OF DETECTION AND USES THEREOF

```



```
; FILE REFERENCE: CL001484
; CURRENT APPLICATION NUMBER: US/60/500,315
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 69978
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11594
; LENGTH: 34227
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-500-315-11594

Query Match      12.5%; Score 28.8; DB 7; Length 34227;
Best Local Similarity 58.0%; Pred. No. 28;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 115 GACAGCGGAGGTGGCCAGACCTCTCTGGGTGATGAGCCTAAGGCGATGTTTGTGTAGG 174
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17633 GAGGCTCGGGCTGCGCAGGAGCCACAGAGGGGGGAGGTAAGGAATGGCAGGCTGCAGG 17692

QY 175 GCCCTATGCTTGCACACTGGGGATCAG 202
    ||| ||| ||||| ||||| |||||
Db 17693 TCCCAAGCCCTGCCCCACAGGGAGGCAG 17720

RESULT 15
US-60-487-610-20089/c
; Sequence 20089, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20089
; LENGTH: 25821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-20089

Query Match      12.3%; Score 28.4; DB 7; Length 25821;
Best Local Similarity 58.1%; Pred. No. 35;
Matches 50; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 51 CTGTGGAGGCACGCTCTGATTGCATGAAGGTCAGTGTCTCTAGTTCCTTCCCCCAGGAA 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 22839 CTAGGGAATGCTGGTTCACATCGGTGGAAGGTTAGGGTTCGGGGGAAGCTGGCAGCAGCA 22780

QY 111 AAACACACGGGAGCTGGCCAAAGCC 136
    ||||| ||||| ||||| ||||| |||||
Db 22779 GATGCACACCGGACCTGGTCAAGGCC 22754

Search completed: October 9, 2003, 23:37:22
Job time : 179 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 . Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 20:13:18 ; Search time 256 Seconds
(without alignments)
2435.821 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 ggggggctggagctaaagca.....ttcaccctagggtgtgctt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231	100.0	231	21	AAZ55983
2	231	100.0	393	21	AAZ55986
3	224	97.0	226	21	AAZ55987
4	114.4	49.5	75798	25	ABX77212
5	89.4	38.7	1870	25	ABZ80240
6	89.4	38.7	5018	6	AAZ50150
7	89.4	38.7	5018	8	AAZ70646
8	87.8	38.0	3336	6	AAZ50149
					Core fragment B, a murine genomic pos
					Murine genomic pos
					Post-transcription
					Mouse uridine diph
					Mouse tramdorin 3
					Sequence of the cd
					Clone 23Bp102 enc
					Sequence of the cd

9	86.2	37.3	1434	18	AAV04900
10	82	35.5	1434	25	ABX11082
11	81.4	35.2	1400	18	AAV04901
12	78.6	34.0	1400	25	ABX11083
13	35	15.2	50	16	AAQ83497
14	34	14.7	122186	22	AAZ89560
15	30.4	13.2	475	24	ABL81597
16	30.4	13.2	691	21	AAZ01644
17	30	13.0	30	21	AAZ55984
18	30	13.0	610	20	AAZ61423
19	29.4	12.7	32187	22	AAZ32249
20	29.4	12.7	32187	24	ABZ67552
21	29.2	12.6	540	20	AAZ08623
22	29.2	12.6	540	21	AAZ35795
23	29.2	12.6	1111	22	AAZ94634
24	29.2	12.6	2648	25	ACC46566
25	29.2	12.6	3881	22	AAI69907
26	29.2	12.6	5176	25	ACC46562
27	29.2	12.6	9040	24	ABL40559
28	29.2	12.6	9154	24	ABZ71695
29	29.2	12.6	9418	24	ABZ71691
30	29.2	12.6	9894	24	ABZ71691
31	29.2	12.6	9993	24	ABZ71690
32	29.2	12.6	9993	24	ABZ71692
33	29.2	12.6	14834	24	ABZ83570
34	29	12.6	820	21	AAZ51239
35	29	12.6	823	21	AAZ35135
36	28.8	12.5	663	22	AAZ33056
37	28.8	12.5	664	22	AAZ33280
38	28.8	12.5	703	21	AAZ01665
39	28.6	12.4	427	23	ABZ42209
40	28.6	12.4	2529	21	AAZ29467
41	28.6	12.4	2588	21	AAZ29469
42	28.6	12.4	2692	14	AAQ51114
43	28.6	12.4	3721	23	ABL07213
44	28.6	12.4	6465	23	ABL07212
45	28.6	12.4	6549	24	AAZ94790

ALIGNMENTS

RESULT 1

AAZ55983
ID AAZ55983 standard; DNA; 231 BP.
XX
AC AAZ55983;
XX
DT 10-APR-2000 (first entry)
XX
DE Core fragment B, a murine genomic core PRE region.
XX
KW Post translational regulatory element; PRE; murine; NCTE;
KW post-transcriptional RNA nucleo-cytoplasmic transport element;
KW attenuated virus; retrovirus; HIV-1; intracisternal A-type particle; IAP;
KW vaccine; viral infection; core fragment B; ds.
XX
OS Mus sp.
XX
PN WO9961596-A2.
XX
PD 02-DEC-1999.
XX
PF 18-MAY-1999; 99WO-US11082.
XX
PR 22-MAY-1998; 98US-0086487.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pavlakis GN, Nappi F;
XX
DR WPI; 2000-116336/10.
XX

Nucleotide sequenc
DNA sequence of ne
Nucleotide sequenc
DNA sequence of ne
DNA probe 3 detect
Human histone deac
Human ovarian canc
Human colon cancer
Murine post transl
DNA encoding a hum
Human DNA repair a
Novel human DNA re
EST with homology
Expressed sequence
Human full-length
Human dithp secret
Human fibrillin 3
Human dithp secret
Human fibrillin 3
DNA encoding human
DNA encoding human
DNA encoding human
Human cDNA differe
Arabidopsis thalia
Arabidopsis thalia
DNA encoding human
DNA encoding human
Human colon cancer
Genomic sequence #
P. chrysogenum sut
Genomic P. chrysog
Sequence encoding
AAQ51114
Drosophila melanog
Drosophila melanog
Human DNA sequenc

PT Novel post-transcriptional regulatory elements used to construct
 XX attenuated retroviruses for vaccines
 PS Claim 3: Page 44; 59pp; English.
 XX This sequence represents a core post-transcriptional regulatory
 CC element (PRE) region, designated core fragment B, which is located
 CC within PRE7 (AAZ55986). Core fragment B can function as a post-
 CC transcriptional RNA nucleocytoplasmic transport element (NCTE). The PRE
 CC is used to replace the NCTE of HIV-1, producing a virus with lower
 CC replicative activity, resulting in an attenuated virus. This novel PRE
 CC was initially derived from a murine genomic sequence and sequence
 CC analysis found that it had significant homology to intracisternal A-type
 CC particle (IAP) sequences. The nucleic acids and oligonucleotides of the
 CC invention can be delivered into cells cultures, tissues and organisms
 CC for synthesis, mutation and screening. When the PRE is used to replace
 CC the NCTE of viruses, especially retroviruses such as HIV, an attenuated
 CC virus is produced, which may be used in a viral vaccine for the
 CC prophylaxis or amelioration of a viral infection in a mammal.
 XX Sequence 231 BP; 47 A; 58 C; 70 G; 56 T; 0 other;
 SQ Query Match 100.0%; Score 231; DB 21; Length 231;
 Best Local Similarity 100.0%; Pred. No. 3e-70;
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTGTGTTGGCATCTCTGGAAGG 60
 DB 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTGTGTTGGCATCTCTGGAAGG 60
 QY 61 CAGCTCTGATGTCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
 DB 61 CAGCTCTGATGTCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
 QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
 DB 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
 QY 181 ATGCTTGACACACACCTGAGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231
 DB 181 ATGCTTGACACACCTGAGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231
 RESULT 2
 AAZ55986
 ID AAZ55986 standard; DNA; 393 BP.
 AC AAZ55986;
 XX
 XX 10-APR-2000 (first entry)
 XX Murine genomic post-transcriptional regulatory element, PRE7.
 DE
 XX Post translational regulatory element; PRE; murine; NCTE;
 KW post-transcriptional RNA nucleocytoplasmic transport element;
 KW attenuated virus; retrovirus; HIV-1; Intracisternal A-type particle; IAP;
 KW vaccine; viral infection; PRE7; ds.
 XX Mus sp.
 XX WO9961596-A2.
 XX
 XX 02-DEC-1999.
 XX
 XX 18-MAY-1999; 99WO-US11082.
 XX
 XX 22-MAY-1998; 98US-0086487.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Pavlakis GN, Nappi F;
 XX WPI; 2000-116336/10.
 DR

XX Novel post-transcriptional regulatory elements used to construct
 PT attenuated retroviruses for vaccines
 PS Example 1: Page 44; 59pp; English.
 XX The invention relates to the core region of a novel post-transcriptional
 CC regulatory element (PRE), designated core fragment B (AAZ55983), that
 CC can function as a post-transcriptional RNA nucleocytoplasmic transport
 CC element (NCTE). The PRE is used to replace the NCTE of HIV-1, producing
 CC a virus with lower replicative activity, resulting in an attenuated
 CC virus. This novel PRE was initially derived from a murine genomic
 CC sequence and sequence analysis found that it had significant homology to
 CC intracisternal A-type particle (IAP) sequences. The nucleic acids and
 CC oligonucleotides of the invention can be delivered into cells cultures,
 CC tissues and organisms for synthesis, mutation and screening. When the
 CC PRE is used to replace the NCTE of viruses, especially retroviruses such
 CC as HIV, an attenuated virus is produced, which may be used in a viral
 CC vaccine for the prophylaxis or amelioration of a viral infection in a
 CC mammal. This sequence represents PRE7, a murine sequence which comprises
 CC core fragment B which was characterised in an exemplification of the
 CC present invention, and is functional as a PRE.
 XX Sequence 393 BP; 85 A; 97 C; 109 G; 102 T; 0 other;
 SQ Query Match 100.0%; Score 231; DB 21; Length 393;
 Best Local Similarity 100.0%; Pred. No. 3.6e-70;
 Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTGTGTTGGCATCTCTGGAAGG 60
 DB 99 GTGGGTGCGAGGCTAAGCACTGACAGAGGATAGCTGTGTTGGCATCTCTGGAAGG 158
 QY 61 CAGCTCTGATGTCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 120
 DB 159 CAGCTCTGATGTCATGAAGTTTCAGTGTCTCTAGTTCCTTCCCTCCAGGAAAAACGACAG 218
 QY 121 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
 DB 219 GGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 278
 QY 181 ATGCTTGACACACCTGAGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 231
 DB 279 ATGCTTGACACACCTGAGGATCAGACCTCTACCTTCACCCATGAGGCTTGCCT 329
 RESULT 3
 AAZ55987
 ID AAZ55987 standard; DNA; 226 BP.
 AC AAZ55987;
 XX
 XX 10-APR-2000 (first entry)
 XX Post-transcriptional regulatory element PRE7 functional fragment, M4.
 DE
 XX Post translational regulatory element; PRE; murine; NCTE;
 KW post-transcriptional RNA nucleocytoplasmic transport element;
 KW attenuated virus; retrovirus; HIV-1; Intracisternal A-type particle; IAP;
 KW vaccine; viral infection; M4; ds.
 XX Unidentified.
 XX WO9961596-A2.
 XX
 XX 02-DEC-1999.
 XX
 XX 18-MAY-1999; 99WO-US11082.
 XX
 XX 22-MAY-1998; 98US-0086487.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX

PI Pavlakis GN, Nappi F;
 XX WPI; 2000-116336/10.
 DR
 XX
 PT Novel post-transcriptional regulatory elements used to construct
 PT attenuated retroviruses for vaccines -
 XX
 PS Example 1; Page 45; 59pp; English.
 XX
 CC The invention relates to the core region of a novel post-transcriptional
 CC regulatory element (PRE), designated core fragment B (AA255983), that
 CC can function as a post-transcriptional RNA nucleocytoplasmic transport
 CC element (NCTE). The PRE is used to replace the NCTE of HIV-1, producing
 CC a virus with lower replicative activity, resulting in an attenuated
 CC virus. This novel PRE was initially derived from a murine genomic
 CC sequence and sequence analysis found that it had significant homology to
 CC intracellular A-type particle (IAP) sequences. The nucleic acids and
 CC oligonucleotides of the invention can be delivered into cells cultures,
 CC tissues and organisms for synthesis, mutation and screening. When the
 CC PRE is used to replace the NCTE of viruses, especially retroviruses such
 CC as HIV, an attenuated virus is produced, which may be used in a viral
 CC vaccine for the prophylaxis or amelioration of a viral infection in a
 CC mammal. This sequence represents M4, a sequence which is very similar to
 CC core fragment B (having 7 additional nucleotides at the 3' end and 2
 CC fewer nucleotides at the 5' end). M4 was characterised in an
 CC exemplification of the present invention, and is functional as a PRE.
 XX
 SQ Sequence 226 BP; 47 A; 58 C; 69 G; 52 T; 0 other;
 Query Match 97.0%; Score 224; DB 21; Length 226;
 Best Local Similarity 100.0%; Pred. No. 8e-68;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTGGGTGCGAGCTAAGCACTGCACAGAGGATGCTGTGGCATCTCTGGAAGG 60
 DB 3 GTGGGTGCGAGCTAAGCACTGCACAGAGGATGCTGTGGCATCTCTGGAAGG 62
 QY 61 CAGGTCTGATGTCATGAAGTTCAGTGTCTAGTTCCTTCCCTCCAGGAACGACACG 120
 DB 63 CAGGTCTGATGTCATGAAGTTCAGTGTCTAGTTCCTTCCCTCCAGGAACGACACG 122
 QY 121 GGAGCTGGCCCAAGCACTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTAGGCCCCCT 180
 DB 123 GGAGCTGGCCCAAGCACTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTAGGCCCCCT 182
 QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCCATGAGG 224
 DB 183 ATGCTTGCACACTGGGGATCAGACCTCTACCTTACCCCATGAGG 226
 RESULT 4
 ABX77212
 ID ABX77212 standard; DNA; 75798 BP.
 AC
 XX
 AC ABX77212;
 XX
 DT 01-MAY-2003 (first entry)
 XX
 DE Mouse uridine diphosphoglucuronosyl transferase gene locus.
 DE
 KW Human; ss; transgenic; drug metabolism; behaviour; PCR; primer; mouse;
 KW pharmacokinetic assay; pharmacodynamic assay; toxicology; serum albumin;
 KW alpha-acidic glycoprotein; CYP; multidrug resistance protein; MRP;
 KW uridine diphosphoglucuronosyl transferase; UGT; cytochrome P450.
 XX
 OS Homo sapiens.
 XX
 PN W0200283897-A1.
 XX
 XX 24-OCT-2002.
 PD
 XX 18-APR-2002; 2002WO-AU00485.
 PF
 XX

PR 18-APR-2001; 2001AU-0004467.
 XX (GENE-) GENE STREAM PTY LTD.
 PA
 XX Dally JM;
 XX
 DR WPI; 2003-093021/08.
 XX
 PT New transgenic non-human animal expressing a foreign polypeptide
 PT associated with drug behavior and/or metabolism, useful for studying
 PT the behavior and/or metabolism of a drug in other animals -
 XX
 PS Disclosure; Page 354-394; 408pp; English.
 XX
 CC This invention relates to a transgenic non-human animal which may be
 CC used for assessing the behaviour and/or metabolism of a drug in another
 CC animal and which expresses a foreign polypeptide associated with drug
 CC behaviour and/or metabolism. The invention also comprises a nucleic acid
 CC construct for use in producing the above transgenic non-human animal
 CC and a method of assessing the metabolism and/or behavior of a drug in
 CC an animal of interest, comprising administering a test agent to the
 CC transgenic animal and conducting analytical tests to determine drug
 CC metabolism and/or behaviour. The transgenic animal is useful in studying
 CC drug metabolism and/or behaviour in other animals. The nucleic acid
 CC construct is useful in producing the above transgenic animal and the
 CC methods are used for producing, breeding and using transgenic animals
 CC for pharmacological (e.g. pharmacokinetic or pharmacodynamic assays)
 CC and/or toxicological studies. Nucleic acid sequences used within
 CC the invention are serum albumin; alpha-acidic glycoprotein; cytochrome
 CC P450 (CYP); uridine diphosphoglucuronosyl transferase (UGT); multidrug
 CC resistance proteins and (MRP's). The present sequence represents a
 CC PCR primer used to create a transgenic animal within the scope of the
 CC invention.
 XX
 SQ Sequence 75798 BP; 21534 A; 17028 C; 16619 G; 20617 T; 0 other;
 Query Match 49.5%; Score 114.4; DB 25; Length 75798;
 Best Local Similarity 81.1%; Pred. No. 8.6e-29;
 Matches 133; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 55 GGAAGCAGCTCTGATTGATGAAGTTCAGTGTCTAGTTCCTTCCCTCCAGGAAC 114
 DB 75206 GGAAGCAGCTCTGATTGATGAAGTTCAGTGTCTAGTTCCTTCCCTCCAGGAAC 114
 QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGATGAGCCCTAAGGGATGTTTGTAGG 174
 DB 75266 GACACGGGAGCTGGCCCAAGACCTCTCGGGTGAGAGCTGGGAGAGGATTTGTGTAAT 75325
 QY 175 GCGCCCTATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
 DB 75326 GCGCCCTATGCTTGCACACTGGGGATTTGACCTCTATCTCCACTC 75369
 RESULT 5
 ABZ80240/C
 ID ABZ80240 standard; cDNA; 1870 BP.
 XX
 AC ABZ80240;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Mouse tramdorin 3' RACE product encoding cDNA SEQ ID NO:40.
 DE
 KW Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
 KW central nervous system disorder; CNS disorder; multiple sclerosis;
 KW nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramd;
 KW tramdorin; mouse; tramdorin 3; gene; ss.
 XX
 OS Mus sp.
 XX
 PH Key Location/Qualifiers
 PH CDS 11..970
 FT /*tag= a

FT /partial
 FT /product= "tramdorin 3 fragment"
 FT /note= "no start codon given"
 PN WO2003016502-A2.
 XX 27-FEB-2003.
 XX
 PF 21-AUG-2002; 2002WO-US26637.
 XX
 PR 21-AUG-2001; 2001US-313907P.
 PR 21-AUG-2002; 2002US-0225810.
 XX
 PA (MCLA-) MCLAUGHLIN RES INST.
 XX
 PI Birmingham JR;
 XX
 DR WPI: 2003-278567/27.
 DR P-PSDB: ABP96445.
 XX
 PT New nucleic acid sequence encoding tramdorins, e.g. mouse tramd 1,
 PT mouse tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd
 PT 3 or rat tramd 1, useful for treating CNS, e.g. stroke, multiple
 PT sclerosis, trauma, neuropathic pain
 XX
 PS Example 11; Fig 23C; 177pp; English.
 XX
 CC The present invention describes an isolated nucleic acid sequence
 CC comprising a cDNA sequence encoding mouse tramdorin (tramd) 2, mouse
 CC tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or
 CC the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is
 CC located to chromosome 11, whereas human tramd 1 is located to chromosome
 CC 5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic
 CC and cerebroprotective activities, and can be used in gene therapy. The
 CC nucleic acid sequences are useful for diagnosing and treating central
 CC nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
 CC neuropathic pain, stroke or trauma, and non-CNS disorders. The present
 CC sequence encodes a mouse tramd 3' RACE product, which is given in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 1870 BP; 483 A; 484 C; 429 G; 474 T; 0 other;
 Query Match 38.7%; Score 89.4; DB 25; Length 1870;
 Best Local Similarity 74.9%; Pred. No. 1.1e-20;
 Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;
 QY 55 GGAAGGACGCTCTGATTGCATGAAGGTTCACTGTCCTAGTTCCTCCCGCAGGAAAAAC 114
 DB 1347 GGGAGACATGTCATCTTTCATGAAGGTTCACTGTCCTAGTTCCTCCCGCAGGAAAAAC 1288
 QY 115 GACACGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTT 167
 DB 1287 GACACGGAGCAGGTCAGGTTGCTCTGGGTAAAAAGCCTGTAAGCCTAAGAGCTAATCCT 1228
 QY 168 GTGTAGGCCCCCTATGCTTCACACTGGGATCAGACCTCTACCTTTCACCC 218
 DB 1227 GTACATGGCTCTTACCTACACACTGGGATTTGACCTCTATCTCCATC 1177
 RESULT 6
 AAN50150
 ID AAN50150 standard; cDNA: 5018 BP.
 XX
 AC AAN50150;
 XX
 XX 25-MAR-2003 (updated)
 DT 16-AUG-2002 (updated)
 DT 27-SEP-1991 (first entry)
 XX
 DE Sequence of the cDNA clone 23B6p10.2 encoding a polypeptide
 DE exhibiting mammalian immunoglobulin binding factor activity (IBF).
 XX
 KW Immunoglobulin E-mediated disease; therapy; B-cell differentiation;

KW immunoglobulin E-binding factor; ss.
 XX Rattus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 423..2750
 FT /*tag= a
 XX
 PN EPI55192-A.
 XX
 PD 18-SEP-1985.
 XX
 PF 15-MAR-1985; 85EP-0301834.
 XX
 PR 16-MAR-1984; 84US-0590430.
 XX
 PA (SCHE) SCHERING BIOTECH CORP.
 PA (UYJO) UNIV JOHNS HOPKINS.
 PA (SCHE) SCHERING-BIOTECH CO.
 PA (DNAX-) DNAX RES INST MOLEC.
 XX
 PI Martens CL, Ishizaka K, Moore KW, Huff TF;
 XX
 DR WPI: 1985-231863/38.
 DR P-PSDB: AAP50121.
 XX
 PT New complementary DNA clones coding for poly:peptide(s) - with
 PT sequence of mammalian immunoglobulin factor and obtd. from
 PT transformed or transfected host
 XX
 PS Claim 8; Page 59-64; 71pp; English.
 XX
 CC IBF and IBF cDNA are useful in studies on the immune system.
 CC Treatment for IgE-mediated diseases may be possible. IBF may enhance
 CC B-cell differentiation into an immunoglobulin-secreting cell.
 CC (Updated on 16-AUG-2002 to add missing OS field.)
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 5018 BP; 1424 A; 1086 C; 1165 G; 1343 T; 0 other;
 Query Match 38.7%; Score 89.4; DB 6; Length 5018;
 Best Local Similarity 74.9%; Pred. No. 1.6e-20;
 Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;
 QY 55 GGAAGGACGCTCTGATTGCATGAAGGTTCACTGTCCTAGTTCCTCCCGCAGGAAAAAC 114
 DB 4449 GGGAGACATGTCATCTTTCATGAAGGTTCACTGTCCTAGTTCCTCCCGCAGGAAAAAC 4508
 QY 115 GACACGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTT 167
 DB 4509 GACACGGAGCAGGTCAGGTTGCTCTGGGTAAAAAGCCTGTAAGCCTAAGAGCTAATCCT 4568
 QY 168 GTGTAGGCCCCCTATGCTTCACACTGGGATCAGACCTCTACCTTTCACCC 218
 DB 4569 GTACATGGCTCTTACCTACACACTGGGATTTGACCTCTATCTCCATC 4619
 RESULT 7
 AAN70646
 ID AAN70646 standard; cDNA: 5018 BP.
 XX
 AC AAN70646;
 XX
 XX 25-MAR-2003 (updated)
 DT 10-MAR-2003 (updated)
 DT 27-APR-1991 (first entry)
 XX
 DE Clone 23B6p102 encoding polypeptide with IgE binding factor
 DE activity.
 XX
 KW Antibody; immunoglobulin; ss.
 XX
 OS Unidentified.

```

XX FH Key Location/Qualifiers
XX FT CDS 423..2753
XX FT /*tag= a
XX PN JP62045600-A.
XX PD 27-FEB-1987.
XX PF 21-AUG-1985; 85JP-0183810.
XX PR 21-AUG-1985; 85JP-0183810.
XX PA (SCHE ) SCHERING BIOTECH CORP.
XX DR WPI; 1987-097765/14.
XX DR P-PSDB; AAP70417.
XX PT cDNA clone coding polypeptide - showing IgE bond factor activity.
XX PS Disclosure; Page 29; 33pp; Japanese.
XX CC The clone encodes a polypeptide which shows IgE binding factor
XX CC activity. See also AAN70645.
XX CC (Updated on 10-MAR-2003 to add missing OS field.)
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 5018 BP; 1423 A; 1087 C; 1165 G; 1343 T; 0 other;
Query Match 38.7%; Score 89.4; DB 8; Length 5018;
Best Local Similarity 74.9%; Pred. No. 1.6e-20;
Matches 128; Conservative 0; Mismatches 36; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTGCATGAAGGTTTCAGTGTCTTCCCTTCCCCAGGAAAAAC 114
DB 4449 GGGAGACATGTCATCTTTCATGAAGTTTCAGTGTCTTCCCTTCCCCAGGAAAAAC 4508
QY 115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTT 167
DB 4509 GACACGGGAGCAGGTCTGGGTGCTCTGGGTAAAGCCTGTGAAGCCTAAGAGCTAATCCT 4568
QY 168 GTGTAGGCCCTTATGCTTGCACACTGGGGATCAGACCTCTACCTTACACC 218
DB 4569 GTACATGGCTCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 4619

RESULT 8
AAN50149
ID AAN50149 standard; cDNA; 3336 BP.
XX AC AAN50149;
XX DT 25-MAR-2003 (updated)
XX DT 16-AUG-2002 (updated)
XX DT 27-SEP-1991 (first entry)
XX DE Sequence of the cDNA clone 23b6p8.3 encoding a polypeptide
XX DE exhibiting mammalian immunoglobulin binding factor activity (IBF).
XX KW Immunoglobulin E-mediated disease; therapy; B-cell differentiation;
XX KW immunoglobulin E-binding factor; ss.
XX OS Rattus sp.
XX PN
XX FH Key Location/Qualifiers
XX FT CDS 94..1767
XX FT /*tag= a
XX PN EP155192-A.
XX PD 18-SEP-1985.
XX PF 15-MAR-1985; 85EP-0301834.

```

```

XX PR 16-MAR-1984; 84US-0590430.
XX PA (SCHE ) SCHERING BIOTECH CORP.
XX PA (UYJO ) UNIV JOHNS HOPKINS.
XX PA (SCHE ) SCHERING-BIOTECH CO.
XX PA (DNAX-) DNAX RES INST MOLEC.
XX PI Martens CL, Ishizaka K, Moore KW, Huff TF;
XX WPI; 1985-231863/38.
XX DR P-PSDB; AAP50120.
XX PT New complementary DNA clones coding for poly:peptide(s) - with
XX PT sequence of mammalian immunoglobulin factor and obtd. from
XX PT transformed or transfected host
XX PS Claim 8; Page 55-58; 71pp; English.
XX CC IBF and IBF cDNA are useful in studies on the immune system.
XX CC Treatment for IgE-mediated diseases may be possible. IBF may enhance
XX CC B-cell differentiation into an immunoglobulin-secreting cell.
XX CC (Updated on 16-AUG-2002 to add missing OS field.)
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 3336 BP; 944 A; 725 C; 839 G; 828 T; 0 other;
Query Match 38.0%; Score 87.8; DB 6; Length 3336;
Best Local Similarity 74.3%; Pred. No. 5.1e-20;
Matches 127; Conservative 0; Mismatches 37; Indels 7; Gaps 1;
QY 55 GGAAGGCACGCTGATTGCATGAAGGTTTCAGTGTCTTCCCTTCCCCAGGAAAAAC 114
DB 2772 GGGAAACATGTCATCTTTCATGAAGTTTCAGTGTCTTCCCTTCCCCAGGAAAAAC 2831
QY 115 GACACGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCTAAGGGATGGTTT 167
DB 2832 GACACGGGAGCAGGTCTGGGTGCTCTGGGTAAAGCCTGTGAAGCCTAAGAGCTAATCCT 2891
QY 168 GTGTAGGCCCTTATGCTTGCACACTGGGGATCAGACCTCTACCTTACACC 218
DB 2892 GTACATGGCTCTTTACCTACACACTGGGGATTGACCTCTATCTCCACTC 2942

RESULT 9
AAV04900
ID AAV04900 standard; DNA; 1434 BP.
XX AC AAV04900;
XX DT 30-APR-1998 (first entry)
XX DE Nucleotide sequence of a portion of a neo-minichromosome.
XX KW Neo-minichromosome; multicentric; dicentric; chromosome; euchromatin;
XX KW heterochromatic DNA; minichromosome; artificial chromosome;
XX KW selectable marker; satellite artificial chromosome; SATAC; genomic locus;
XX KW targeted integration; transgenic animal; therapeutic product;
XX KW gene therapy; cloning vehicle; genomic DNA library; ss.
XX OS Mus sp.
XX PN
XX PN WO9740183-A2.
XX PD 30-OCT-1997.
XX PF 10-APR-1997; 97WO-US05911.
XX PR 07-AUG-1996; 96US-0695191.
XX PR 10-APR-1996; 96US-0629822.
XX PR 15-JUL-1996; 96US-0682080.
XX PA (AMGE-) AMERICAN GENE THERAPY INC.

```

[illegible]

KW XX bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; mutant; ds.
 OS XX Homo sapiens.
 OS Mus sp.
 OS Bacteriophage lambda.
 OS Bacteriaceae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..1434
 FT 1..1434
 FT /*tag=
 FT /note= "All N nucleotides have been added by the
 FT indexer to match the number of bases this
 FT sequence is listed to contain in the
 FT sequence listing"
 XX
 PN US2002160410-A1.
 XX
 PD 31-OCT-2002.
 XX
 PF 17-APR-2002; 2002US-0125767..
 XX
 PF 28-NOV-2000; 2000US-0724693.
 PR 10-APR-1997; 97US-0835682.
 PR 10-APR-1996; 96US-0629822.
 PR 15-JUL-1996; 96US-0682080.
 PR 07-AUG-1996; 96US-0695191.
 XX
 PA (HADL/) HADLACZY G.
 PA (SZAL/) SZALAY A A.
 XX
 FI Hadlaczky G, Szalay AA;
 PI WPI; 2003-265757/26..
 DR
 XX
 PT Amplifying nucleic acid for constructing artificial chromosomes,
 PT comprises introducing a DNA fragment into a cell to incorporate the DNA
 PT into a chromosome, and identifying cells with chromosomes that has
 PT undergone amplification
 XX
 PS Example 2; Page 63; 107pp; English.
 XX
 CC The present invention relates to methods for amplifying nucleic acids
 CC in cells, for isolating artificial chromosomes, and preparing cell
 CC lines that contain artificial chromosomes. The methods comprise
 CC introducing a DNA fragment comprising a selectable marker into a cell,
 CC growing the cell under selective conditions to produce cells that have
 CC incorporated the DNA fragment or its portion into a chromosome, and
 CC identifying from among the resulting cells those that include a
 CC chromosome or its fragment with a portion that has undergone
 CC amplification. The invention also discloses cells and cell lines that
 CC contain the nucleic acids or artificial chromosomes. The methods of
 CC the invention are useful for amplifying nucleic acids in cells, in
 CC generating and isolating artificial chromosomes (e.g. mammalian
 CC artificial chromosomes or MACs), and in delivering the chromosomes
 CC to selected cells and tissues. The artificial chromosomes are useful
 CC in gene therapy, gene product production systems, production
 CC of humanised genetically transformed animal organs, production of
 CC transgenic plants and animals that would employ chromosomal elements as
 CC information storage vehicles, for analysis and study of centromere
 CC function, for the production of artificial chromosome vectors, and for
 CC the preparation of species-specific artificial chromosomes. The present
 CC sequence represents a neo-minichromosome (composed of human,
 CC bacteriophage lambda, plasmid, bacterial and mouse DNAs) PCR product
 CC from EC3/7C5 cells that is used in the method of the present invention.
 XX
 SQ Sequence 1434 BP; 336 A; 310 C; 333 G; 403 T; 52 other;

Db 904 GACATGTCATCTTCAATGAAGGTTCAAGTCTCCTAGTTCCTCCCGGCAAAANNACA 963
 QY 119 CGGGAGCTGCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAGGGATGGTTTGTGT 171
 Db 964 CGGGAGCAGGTCAAGGTTGCTCTGGGTAAAGCCTGTGAGCCTGGGAGCTAATCANNATC 1023
 QY 172 AGGGCCCCATGTTGCACACTGGGGATCAGACCTCTACCTTCACCC 218
 Db 1024 ATGGCTCTTTACCTACACTAGGGGATTTGACCTCTATCTCCACTC 1070
 RESULT 11
 ID AAV04901 standard; DNA; 1400 BP.
 AC AAV04901;
 XX
 DT 30-APR-1998 (first entry)
 XX
 DE Nucleotide sequence of a portion of a neo-minichromosome.
 XX
 KW Neo-minichromosome; multicentric; dicentric; chromosome; euchromatin;
 KW heterochromatic DNA; minichromosome; artificial chromosome;
 KW selectable marker; satellite artificial chromosome; SATAC; genomic locus;
 KW targeted integration; transgenic animal; therapeutic product;
 KW gene therapy; cloning vehicle; genomic DNA library; ss.
 OS Mus sp.
 XX
 PN W09740183-A2.
 XX
 PD 30-OCT-1997.
 XX
 PF 10-APR-1997; 97WO-0505911.
 XX
 PR 07-AUG-1996; 96US-0695191.
 PR 10-APR-1996; 96US-0629822.
 PR 15-JUL-1996; 96US-0682080.
 XX
 PA (AMGE-) AMERICAN GENE THERAPY INC.
 PA (BIOL-) BIOLOGICAL RES CENT HUNGARIAN ACAD SCI.
 PA (UYLO-) UNIV LOMA LINDA.
 XX
 PI Hadlaczky G, Szalay AA;
 XX
 DR WPI; 1997-535860/49.
 XX
 PT Producing satellite artificial chromosomes or mini.chromosomes -
 PT useful for, e.g. cloning multiple proteins of a metabolic pathway or
 PT multivalent vaccines, etc.
 XX
 PS Claim 85; Page 204; 248pp; English.
 XX
 CC Nucleotide sequences AAV04900-902 are portions of a neo-minichromosome
 CC isolated from a mouse rRNA gene. 3 products were obtained from a single
 CC amplification reaction, which suggests that the sequence of the DNA
 CC located between different sets of inverted repeats may differ.
 CC AAV04900-01 show high (96%) sequence homology to portions of DNA from
 CC intracisternal A-particles from mouse. The minichromosome is derived from
 CC multicentric, typically dicentric, chromosome that contains more
 CC euchromatin than heterochromatic DNA. The minichromosome is an example
 CC of an artificial chromosome. Artificial chromosomes can be produced by
 CC incorporating a DNA fragment comprising a selectable marker into a cells'
 CC genomic DNA, into or adjacent to an amplifiable region, and selecting a
 CC cell that comprises either a satellite artificial chromosome (SATAC)
 CC (see AAV04903-09), a de novo centromere, or a minichromosome (e.g.
 CC present sequence). Artificial chromosomes provide an extra genomic locus
 CC for targeted integration of megabase size DNA fragments that contain
 CC single or multiple genes. SATACs can be introduced into embryonic cells
 CC of non-human animals to produce transgenic animals that express a
 CC heterologous DNA encoding a therapeutic product, e.g. anti-HIV. The DNA
 CC may also encode antigens that induce immunoprotective response against

CC pathogen. These therapeutic products can be used in gene therapy. The
 CC artificial chromosomes are useful as cloning vehicles that accommodate
 CC entire genomes for the preparation of genomic DNA libraries, and also
 CC for the production of proteins which may be involved in a biochemical
 CC pathway or in multivalent vaccines.
 XX
 SQ Sequence 1400 BP; 341 A; 310 C; 345 G; 404 T; 0 other;
 Query Match 35.2%; Score 81.4; DB 18; Length 1400;
 Best Local Similarity 71.9%; Pred. No. 6.3e-18;
 Matches 123; Conservative 0; Mismatches 41; Indels 7; Gaps 1;
 QY 55 GGAAGGCACGTCGTGATTGTCATGAAGTTCAGTGTCTTAGTTCCTTCCCGGCAAAAC 114
 Db 771 GGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCTTCTCTCCAGGCAAAAC 830
 QY 115 GACACGGGAGCTGGCCCAAGACCTCTCTGGGTGA-----TGAGCCCTAAGGGATGGTTT 167
 Db 831 GACACGGGAGCAGGTGCTCTGGGTAAAGCCTGTGAGCCTAAGAGCTAATCT 890
 QY 168 GTGTAGGGCCCTTATGCTTTCACACTGGGGATCAGACCTCTACCTTCACCC 218
 Db 891 GTACATGGCTCTCTTTACCTACACACTGGGGATTTGACCTCTATCTCCACTC 941
 RESULT 12
 ID ABX11083 standard; DNA; 1400 BP.
 XX
 AC ABX11083;
 XX
 DT 24-APR-2003 (first entry)
 XX
 DE DNA sequence of neo-minichromosome PCR product #2 from EC3/7C5 cells.
 XX
 KW Nucleic acid amplification; artificial chromosome isolation; MAC;
 KW DNA delivery; mammalian artificial chromosome; gene therapy; organ;
 KW humanised genetically transformed animal; chromosomal element;
 KW gene product production system; transgenic; centromere function;
 KW information storage vehicle; artificial chromosome vector; human;
 KW species-specific artificial chromosome; mouse; bacteriophage lambda;
 KW bacteria; plasmid; neo-minichromosome; EC3/7C5 cell; mutant; ds.
 XX
 OS Homo sapiens.
 OS Mus sp.
 OS Bacteriophage lambda.
 OS Bacteriaceae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1..1400
 FT /tag= a
 FT /note= "All N nucleotides have been added by the
 FT index to match the number of bases this
 FT sequence is listed to contain in the
 FT sequence listing"
 XX
 US2002160410-A1.
 XX
 PD 31-OCT-2002.
 XX
 PF 17-APR-2002; 2002US-0125767.
 XX
 PR 28-NOV-2000; 2000US-0724693.
 PR 10-APR-1997; 97US-0835682.
 PR 10-APR-1996; 96US-0629822.
 PR 15-JUL-1996; 96US-0682080.
 PR 07-AUG-1996; 96US-0695191.
 XX
 PA (HADL/) HADLACZKY G.
 PA (SZAL/) SZALAY A A.
 XX
 PI Hadlaczky G, Szalay AA;

```
XX DR WPI; 2003-265757/26.
XX
XX Amplifying nucleic acid for constructing artificial chromosomes,
PT comprises introducing a DNA fragment into a cell to incorporate the DNA
PT into a chromosome, and identifying cells with chromosomes that has
PT undergone amplification
XX
XX Example 2; Page 64; 107pp; English..
XX
XX The present invention relates to methods for amplifying nucleic acids
CC in cells, for isolating artificial chromosomes, and preparing cell
CC lines that contain artificial chromosomes. The methods comprise
CC introducing a DNA fragment comprising a selectable marker into a cell,
CC growing the cell under selective conditions to produce cells that have
CC incorporated the DNA fragment or its portion into a chromosome, and
CC identifying from among the resulting cells those that include a
CC chromosome or its fragment with a portion that has undergone
CC amplification. The invention also discloses cells and cell lines that
CC contain the nucleic acids or artificial chromosomes. The methods of
CC the invention are useful for amplifying nucleic acids in cells, in
CC generating and isolating artificial chromosomes (e.g. mammalian
CC artificial chromosomes or MACs), and in delivering the chromosomes
CC to selected cells and tissues. The artificial chromosomes are useful
CC in gene therapy, gene product production systems, production
CC of humanised genetically transformed animal organs, production of
CC transgenic plants and animals that would employ chromosomal elements as
CC information storage vehicles, for analysis and study of centromere
CC function, for the production of artificial chromosome vectors, and for
CC the preparation of species-specific artificial chromosomes. The present
CC sequence represents a neo-minichromosome (composed of human,
CC bacteriophage lambda, plasmid, bacterial and mouse DNAs) PCR product
CC from EC37/7C5 cells that is used in the method of the present invention.
XX
XX Sequence 1400 BP; 330 A; 299 C; 331 G; 388 T; 52 other;
XX
XX Query Match 34.0%; Score 78.6; DB 25; Length 1400;
XX Best Local Similarity 69.6%; Pred. No. 5.9e-17;
XX Matches 119; Conservative 0; Mismatches 45; Indels 7; Gaps 1;
XX
XX QY 55 GGAAGGACGCTGTCATTCGATGAGTTTCAGTGTCTTCCTCCCGAGGAAAAAC 114
XX DB 771 GGGAGACANNCTCATCTTTCAAGAGGTTGAGTGCCAAGTGCTCTCTCCAGGCAAAAC 830
XX
XX QY 115 GACACGGGAGCTGGCCGACGCTCTCTGGGTGA-----TGAGCCTTAAGGATGGTTT 167
XX DB 831 GACACGGGNNCAGGTCAGGGTGTCTGCGTAAAGCCTGTGAGCCTTAAGAGCTAATCCT 890
XX
XX QY 168 GTGTAGGCGCCCTATGCTTGACACTGGGATCAGACCTCTACCTTACCC 218
XX DB 891 GTACATGNNCCTTTTACCTACACACTGGGATTTGACCTCTATCTCCACTC 941
XX
XX RESULT 13
XX AAQ85497
XX ID AAQ85497 standard; DNA; 50 BP.
XX
XX AC AAQ85497;
XX
XX DT 10-JAN-1996 (first entry)
XX
XX DE DNA probe 3 detects DNA-protein complex in immortal cells.
XX
XX KW DNA-protein complex; detection; proliferation; tumour formation;
XX diagnose; malignancy; biopsy; probe; ss.
XX
XX OS Synthetic.
XX
XX PN WO9502701-A1.
XX
XX PD 26-JAN-1995.
XX
XX PF 13-JUL-1994; 94WO-EP02307.
```

```
XX PR 15-JUL-1993; 93DE-4323727.
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
XX PI Abken H, Albert W, Jungfer H, Abken HJ;
XX WPI; 1995-067344/09.
XX
XX New DNA-protein complex characteristic of cells with unlimited
PT proliferation capacity - and its components and derived antibodies,
PT useful in diagnosing malignant tumours.
XX
XX PS Claim 1; Page 51; 106pp; German.
XX
XX The DNA shown is found in human or animal cells that have an unlimited
CC capacity for unlimited cell proliferation or tumour formation. They have
CC no ability to promote immortalisation of the cells, and are usually
CC found in a DNA-protein complex in the cell cytoplasm. The DNA is useful
CC for detection of these complexes, and diagnosis of malignant tumours.
CC Differentiation between cells with unlimited and only transient
CC proliferation is possible. (See also AAQ85493-54)
XX
XX SQ Sequence 50 BP; 13 A; 15 C; 9 G; 13 T; 0 other;
XX
XX Query Match 15.2%; Score 35; DB 16; Length 50;
XX Best Local Similarity 100.0%; Pred. No. 0.027;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 197 GATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
XX DB 1 GATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 35
XX
XX RESULT 14
XX AAC89560/C
XX ID AAC89560 standard; DNA; 122186 BP.
XX
XX AC AAC89560;
XX
XX DT 08-MAR-2001 (first entry)
XX
XX DE Human histone deacetylase HDAC-D coding sequence.
XX
XX KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;
XX HDAC-D; cell cycle; tumourigenesis; cancer; inhibitor; antisense;
XX gene therapy; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200071703-A2.
XX
XX PD 30-NOV-2000.
XX
XX PF 03-MAY-2000; 2000WO-IB01252.
XX
XX PR 03-MAY-1999; 99US-0132287.
XX
XX PA (METH-) METHYLENE INC.
XX
XX PI Macleod AR, Li Z, Besterman JM;
XX WPI; 2001-016407/02.
XX
XX Antisense oligonucleotide that inhibits expression of a histone
PT deacetylase, useful for treating and/or alleviating the symptoms of
PT neoplasia, or for inhibiting neoplastic cell growth in an animal -
XX
XX PS Disclosure; Page 89-125; 125pp; English.
XX
XX The present invention provides inhibitors of histone deacetylase enzymes
CC such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
CC inhibitors may be antisense strands or they may be compounds identified
```

CC by contacting the enzyme with the compound and measuring the resulting
CC enzyme activity. These inhibitors are useful for treating cancers and for
CC identifying which histone deacetylase is involved in a neoplasia.

xx
SQ Sequence 122186 BP; 29016 A; 31077 C; 32425 G; 29668 T; 0 other;
Query Match 14.7%; Score 34; DB 22; Length 122186;
Best Local Similarity 57.5%; Pred. No. 0.95;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 11 AGGCTAAGCACTGCACAGAGATAGCTGTGGTCATCCTGTGGAAGCACGCTCGAT 70
Dd 69843 AGGCACAGCTCGGGACTCACTACTGTGTGCCCATATTACGCTGGGACTTCTTAC 6978
QY 71 TGCATGAAGTTCCAGTGCTCTAGTTCCTTCCTCCCCAGGAAACA 116
Dd 69783 TTGAGCCCCTTTCTCTCCAGTTCTCTTCGCGAGGAGATCCA 69738

RESULT 15

ABL81597/C

ID ABL81597 standard; cDNA; 475 bp.

XX

AC ABL81597;

XX

DT 17-MAY-2002 (first entry)

XX

DE
XX
Human ovarian c

Human: ovaries

XX

OS Homo sapiens.

XX

PN WO200192581-A2.

XX 200
200

FD
YY
06-DEC-2001.

29-MAY-2001: 2001WO-IIS17756

25-MAY-2000; 2000US-207484P.
(CORI-) CORIXA CORP.
Algate PA, Harlocker SL, Jones R;
WFI; 2002-122075/16.
Composition for therapy and diagnosis of ovarian cancer comprising polypeptide of a ovarian tumor polypeptide, polynucleotide encoding polypeptide, antibody specific to polypeptide or T cell expressing polypeptide -
Claim 1; SEQ ID 4575; 489pp; English.
The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of a ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ABL87934, (III) encoding (II) having a sequence (S2), a T cell population of (II), or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff value and thereby detecting ovarian cancer in the patient, where the amount of polynucleotide hybridising to (IV) is detected preferably by polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is useful for stimulating and/or expanding T cells specific for an ovarian tumour protein comprising contacting T cells with (III) or (II). (III) is useful in design and preparation of ribozyme molecules for inhibiting expression of the tumour polypeptides and proteins in tumour cells; and to isolate a full length gene from a

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2003, 21:30:48 ; Search time 1470 Seconds
(without alignments)
6428.656 Million cell updates/sec

Title: US-09-673-716-1

Perfect score: 231

Sequence: 1 gggggtcgaggctaagca.....ttcaccatgaggtgtgtt 231

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_hgt_hum.*

31: em_hgt_inv.*

32: em_hgt_other.*

33: em_hgt_mus.*

34: em_hgt_pln.*

35: em_hgt_rod.*

36: em_hgt_mam.*

37: em_hgt_vrt.*

38: em_sy.*

39: em_hgt_hum.*

40: em_hgt_mus.*

41: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	231	100.0	1086	10	AF250998	AF250998 Mus muscu
2	231	100.0	1086	10	AF250999	AF250999 Mus muscu
3	231	100.0	2701	10	BC020078	BC020078 Mus muscu
4	231	100.0	42511	2	AC100548	AC100548 Mus muscu
5	231	100.0	64042	2	AC101394	AC101394 Mus muscu
6	231	100.0	66393	2	AC100424	AC100424 Mus muscu
7	231	100.0	66489	2	AC101278	AC101278 Mus muscu
8	231	100.0	66489	2	AC101278	AC101278 Mus muscu
9	231	100.0	70360	2	AC101122	AC101122 Mus muscu
10	231	100.0	71109	10	AL935271	AL935271 Mouse DNA
11	231	100.0	77127	10	BX294125	BX294125 Mouse DNA
12	231	100.0	84250	10	AF481949	AF481949 Mus muscu
13	231	100.0	98653	10	AL731664	AL731664 Mouse DNA
14	231	100.0	110000	2	AC132401	AC132401 Mus muscu
15	231	100.0	110000	2	AL929223_2	Continuation (3 of
16	231	100.0	132457	2	AC102190	AC102190 Mus muscu
17	231	100.0	134025	2	AC107697	AC107697 Mus muscu
18	231	100.0	134264	10	BX005191	BX005191 Mouse DNA
19	231	100.0	134717	10	AL626805	AL626805 Mouse DNA
20	231	100.0	141025	10	AL807250	AL807250 Mouse DNA
21	231	100.0	142902	2	AC119234	AC119234 Mus muscu
22	231	100.0	144387	10	AL844603	AL844603 Mouse DNA
23	231	100.0	149172	10	AC127278	AC127278 Mus muscu
24	231	100.0	150017	2	AL773509	AL773509 Mus muscu
25	231	100.0	155071	2	AC137713	AC137713 Mus muscu
26	231	100.0	156256	2	AC140843	AC140843 Mus muscu
27	231	100.0	161580	2	AC118267	AC118267 Mus muscu
28	231	100.0	162681	10	AC122059	AC122059 Mus muscu
29	231	100.0	164949	2	AC119839	AC119839 Mus muscu
30	231	100.0	165222	2	AC121307	AC121307 Mus muscu
31	231	100.0	165764	2	AC131767	AC131767 Mus muscu
32	231	100.0	166764	2	AC115875	AC115875 Mus muscu
33	231	100.0	166777	2	AC124106	AC124106 Mus muscu
34	231	100.0	167744	10	AC125537	AC125537 Mus muscu
35	231	100.0	171826	2	AC119810	AC119810 Mus muscu
36	231	100.0	172533	2	AC120551	AC120551 Mus muscu
37	231	100.0	173417	10	AL607143	AL607143 Mouse DNA
38	231	100.0	175217	10	AC079680	AC079680 Mus muscu
39	231	100.0	175448	2	AC138680	AC138680 Mus muscu
40	231	100.0	175692	2	AC139226	AC139226 Mus muscu
41	231	100.0	176184	2	AC119918	AC119918 Mus muscu
42	231	100.0	178821	10	AC122385	AC122385 Mus muscu
43	231	100.0	179068	10	AL627125	AL627125 Mouse DNA
44	231	100.0	179662	2	AC137557	AC137557 Mus muscu
45	231	100.0	179699	10	BX072541	BX072541 Mouse DNA

ALIGNMENTS

RESULT 1
AF250998
LOCUS AF250998 1086 bp DNA linear ROD 16-MAY-2001
DEFINITION Mus musculus RTE-clone1 RNA transport element sequence.
ACCESSION AF250998
VERSION AF250998.1 GI:14090507
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1086)
REFERENCE Nappi, F., Schneider, R., Zolotukhin, A., Smulevitch, S.,
Michalowski, D., Bear, J., Felber, B.K. and Pavlakis, G.N.
AUTHORS Identification of a novel posttranscriptional regulatory element by

		/organism="Mus musculus"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:10090"	
		/clone="RTE-clone3"	
misc_feature		381..627	
		/note="fragment M1; contains RNA transport element (RTE), posttranscriptional regulatory element"	
BASE COUNT	247 a 268 c 277 g 294 t		
ORIGIN			
Query Match	100.0%; Score 231; DB 10; Length 1086;		
Best Local Similarity	100.0%; Pred. No. 2e-65;		
Matches 231; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 GTGGGTCGAGCGCTAAGCACTGCACAGAGTAGCTTGTCTGGCATCCTGTGGAGG	60	
Db			
393 GTGGGTCGAGCGCTAAGCACTGCACAGAGTAGCTTGTCTGGCATCCTGTGGAGG	45		
QY	61 CACGCTCTGATTCATGAAGGTTTCAGTGCTCCAGTTCCTTCCCCAGAAAAACGACACG	120	
Db			
453 CACGCTCTGATTCATGAAGGTTTCAGTGCTCCAGTTCCTTCCCCAGAAAAACGACACG	51		
QY	121 GGAGCTGCCAAGACTCTCTGGGTGATGAGCCTTAAGGGATGGTTTTGTGTAGGGCCCT	180	
Db			
513 GGAGCTGCCAAGACTCTCTGGGTGATGAGCCTTAAGGGATGGTTTTGTGTAGGGCCCT	57		
QY	181 ATGCTTGACACTGGGGATCAGACCTCTACCTTACCCTATGAGGCTTGCCT	231	
Db			
573 ATGCTTGACACTGGGGATCAGACCTCTACCTTACCCTATGAGGCTTGCCT	623		
RESULT 3			
BC020078			
LOCUS	2701 bp mRNA linear ROD 20-SEP-2000		
DEFINITION	Mus musculus, similar to RIKEN cDNA 1700066C05 gene, clone		
ACCESSION	MGC:28125 IMAGE:3980327, mRNA, complete cds.		
VERSION	BC020078		
KEYWORDS	MGC.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus		
REFERENCE	1 (bases 1 to 2701)		
AUTHORS	Straussberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabbs@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amgb@bcm.tmc.edu		
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 36 Row: e Column: 4 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.			
FEATURES			
1..2701	Location/Qualifiers		


```

* 15735 16412: contig of 678 bp in length
* 16413 16512: gap of 100 bp
* 16513 17218: contig of 706 bp in length
* 17219 18010: contig of 692 bp in length
* 18011 18110: gap of 100 bp
* 18111 18790: contig of 680 bp in length
* 18791 18890: gap of 100 bp
* 18891 19594: contig of 704 bp in length
* 19595 19694: gap of 100 bp
* 19695 20382: contig of 688 bp in length
* 20383 20482: gap of 100 bp
* 20483 21165: contig of 683 bp in length
* 21166 21265: gap of 100 bp
* 21266 21972: contig of 707 bp in length
* 21973 22073: gap of 100 bp
* 22073 22754: contig of 681 bp in length
* 22754 22853: gap of 100 bp
* 22853 23527: contig of 674 bp in length
* 23527 23627: gap of 100 bp
* 23628 24334: contig of 707 bp in length
* 24335 24434: gap of 100 bp
* 24435 25132: contig of 698 bp in length
* 25133 25232: gap of 100 bp
* 25233 25910: contig of 678 bp in length
* 25911 26010: gap of 100 bp
* 26011 26691: contig of 681 bp in length
* 26692 26791: gap of 100 bp
* 26792 27469: contig of 678 bp in length
* 27470 27569: gap of 100 bp
* 27570 28256: contig of 687 bp in length
* 28257 28356: gap of 100 bp
* 28357 29071: contig of 715 bp in length
* 29072 29171: gap of 100 bp
* 29172 29868: contig of 697 bp in length
* 29869 30719: contig of 751 bp in length
* 30720 30819: gap of 100 bp
* 30820 31527: contig of 708 bp in length
* 31528 32310: contig of 683 bp in length
* 32311 32410: gap of 100 bp
* 32411 33099: contig of 689 bp in length
* 33100 33199: gap of 100 bp
* 33200 33877: contig of 678 bp in length
* 33878 34666: contig of 689 bp in length
* 34667 34766: gap of 100 bp
* 34767 35560: contig of 694 bp in length
* 35561 36271: gap of 100 bp
* 36272 36371: contig of 711 bp in length
* 36372 37045: contig of 674 bp in length
* 37046 37145: gap of 100 bp
* 37146 37835: contig of 690 bp in length
* 37836 37935: gap of 100 bp
* 37936 38609: contig of 674 bp in length
* 38610 38709: gap of 100 bp
* 38710 39386: contig of 677 bp in length
* 39387 39486: gap of 100 bp
* 39487 40166: contig of 680 bp in length
* 40167 40266: gap of 100 bp
* 40267 40937: contig of 671 bp in length
* 40938 41038: gap of 100 bp
* 41039 41717: contig of 680 bp in length
* 41718 41817: gap of 100 bp
* 41818 42511: contig of 694 bp in length.

```

FEATURES

source

```

1. .42511
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-152017"

```

```

BASE COUNT 10831 a 7531 c 7714 g 10953 t 5482 others
ORIGIN

```

```

Query Match 100.0%; Score 231; DB 2; Length 42511;
Best Local Similarity 100.0%; Pred. NO. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGTAGCTTGTGTGGCATCTGTGGAAGG 60
|||||
Db 27042 GTGGGTGCGAGGCTAAGCACTGCACAGAGTAGCTTGTGTGGCATCTGTGGAAGG 27101
|||||

```

```

QY 61 CAGCTCTGATGTCATGAAGTTTCAGTCTCCTAGTCTCCCTCCCGAGAAAACACACG 120
|||||
Db 27102 CAGCTCTGATGTCATGAAGTTTCAGTCTCCTAGTCTCCCTCCCGAGAAAACACACG 27161
|||||

```

```

QY 121 GGAGCTGGCCAAAGCCTCTCTGGTGATGAGCCTTAAGGGATGTTTGTAGGCCCTT 180
|||||
Db 27162 GGAGCTGGCCAAAGCCTCTCTGGTGATGAGCCTTAAGGGATGTTTGTAGGCCCTT 27221
|||||

```

```

QY 181 ATGCTTGCCACATGGGGATCAGACCTCTACCTTCACCATGAGGCTTGCCTT 231
|||||
Db 27222 ATGCTTGCCACATGGGGATCAGACCTCTACCTTCACCATGAGGCTTGCCTT 27272
|||||

```

RESULT 5

```

AC101394 64042 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-119F14, LOW-PASS SEQUENCE SAMPLING.
AC101394

```

```

VERSION AC101394.1 GI:17060169
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus (house mouse)

```

ORGANISM

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 64042)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-119F14
Unpublished

```

REFERENCE

```

2 (bases 1 to 64042)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faroo,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehocsky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrin,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

```

Direct Submission

```

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

COMMENT

```

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```


Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L16611
Center clone name: 119_F_14

* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 692: contig of 692 bp in length
* 792: gap of 100 bp
* 1511: contig of 719 bp in length
* 1512: gap of 100 bp
* 1612: contig of 702 bp in length
* 2313: gap of 100 bp
* 2414: contig of 695 bp in length
* 3109: gap of 100 bp
* 3209: contig of 721 bp in length
* 3930: gap of 100 bp
* 4030: contig of 716 bp in length
* 4746: gap of 100 bp
* 4846: contig of 714 bp in length
* 5560: gap of 100 bp
* 5660: contig of 745 bp in length
* 6403: gap of 100 bp
* 6503: contig of 731 bp in length
* 7235: gap of 100 bp
* 7336: contig of 733 bp in length
* 8069: gap of 100 bp
* 8169: contig of 707 bp in length
* 8876: gap of 100 bp
* 8976: contig of 722 bp in length
* 9698: gap of 100 bp
* 9798: contig of 711 bp in length
* 10509: gap of 100 bp
* 10609: contig of 705 bp in length
* 11314: gap of 100 bp
* 11414: contig of 698 bp in length
* 12112: gap of 100 bp
* 12212: contig of 715 bp in length
* 12927: gap of 100 bp
* 13027: contig of 707 bp in length
* 13734: gap of 100 bp
* 13834: contig of 711 bp in length
* 14545: gap of 100 bp
* 14645: contig of 701 bp in length
* 15346: gap of 100 bp
* 15446: contig of 723 bp in length
* 16169: gap of 100 bp
* 16269: contig of 702 bp in length
* 16971: gap of 100 bp
* 17071: contig of 694 bp in length
* 17765: gap of 100 bp
* 17865: contig of 706 bp in length
* 18571: gap of 100 bp
* 18671: contig of 719 bp in length
* 19390: gap of 100 bp
* 19490: contig of 708 bp in length
* 20198: gap of 100 bp
* 20298: contig of 733 bp in length
* 21031: gap of 100 bp
* 21131: contig of 729 bp in length
* 21860: gap of 100 bp
* 21960: contig of 716 bp in length
* 22676: gap of 100 bp
* 22776: contig of 721 bp in length

* 23497: gap of 100 bp
* 23597: contig of 711 bp in length
* 24407: gap of 100 bp
* 25127: contig of 720 bp in length
* 25227: gap of 100 bp
* 25941: contig of 714 bp in length
* 26041: gap of 100 bp
* 26758: contig of 717 bp in length
* 26858: gap of 100 bp
* 27583: contig of 725 bp in length
* 27684: gap of 100 bp
* 28395: contig of 712 bp in length
* 28495: gap of 100 bp
* 29167: contig of 672 bp in length
* 29267: gap of 100 bp
* 29918: contig of 651 bp in length
* 30018: gap of 100 bp
* 30734: contig of 716 bp in length
* 30834: gap of 100 bp
* 31572: contig of 738 bp in length
* 31672: gap of 100 bp
* 32377: contig of 705 bp in length
* 32477: gap of 100 bp
* 33193: contig of 716 bp in length
* 33293: gap of 100 bp
* 33982: contig of 689 bp in length
* 34082: gap of 100 bp
* 34795: contig of 713 bp in length
* 34895: gap of 100 bp
* 35603: contig of 708 bp in length
* 35703: gap of 100 bp
* 36427: contig of 724 bp in length
* 36527: gap of 100 bp
* 37256: contig of 729 bp in length
* 37356: gap of 100 bp
* 38079: contig of 723 bp in length
* 38179: gap of 100 bp
* 38900: contig of 721 bp in length
* 39001: gap of 100 bp
* 39717: contig of 717 bp in length
* 39817: gap of 100 bp
* 40532: contig of 715 bp in length
* 40632: gap of 100 bp
* 41342: contig of 710 bp in length
* 41442: gap of 100 bp
* 42138: contig of 696 bp in length
* 42238: gap of 100 bp
* 42942: contig of 704 bp in length
* 43042: gap of 100 bp
* 43741: contig of 699 bp in length
* 43841: gap of 100 bp
* 44559: contig of 718 bp in length
* 44659: gap of 100 bp
* 45353: contig of 694 bp in length
* 45453: gap of 100 bp
* 46174: contig of 721 bp in length
* 46274: gap of 100 bp
* 46989: contig of 715 bp in length
* 47089: gap of 100 bp
* 47789: contig of 700 bp in length
* 47889: gap of 100 bp
* 48602: contig of 713 bp in length
* 48702: gap of 100 bp
* 49409: contig of 707 bp in length
* 49509: gap of 100 bp
* 50188: contig of 679 bp in length
* 50288: gap of 100 bp
* 51005: contig of 717 bp in length
* 51105: gap of 100 bp
* 51825: contig of 720 bp in length
* 51925: gap of 100 bp
* 52629: contig of 704 bp in length
* 52729: gap of 100 bp

```

* 52730 53448: contig of 719 bp in length
* 53449 53548: gap of 100 bp
* 53549 54259: contig of 711 bp in length
* 54260 54359: gap of 100 bp
* 54360 55074: contig of 715 bp in length
* 55075 55174: gap of 100 bp
* 55175 55883: contig of 709 bp in length
* 55884 55983: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 64042;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGGAGCTAAGCAGCTGACAGAGATAGTCTGTGTCATCCTGTGGAAGG 60
|||||
Db 32664 GTGGGTGGGAGCTAAGCAGCTGACAGAGATAGTCTGTGTCATCCTGTGGAAGG 32723
|||||

QY 61 CAGTCTGATTCATCAAGGTCAGTCTAGTCCCTCCCGAGGAAACACGACG 120
|||||
Db 32724 CAGTCTGATTCATCAAGGTCAGTCTAGTCCCTCCCGAGGAAACACGACG 32783
|||||

QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTAAGGATGCTTTGTAGGGCCCT 180
|||||
Db 32784 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCTAAGGATGCTTTGTAGGGCCCT 32843
|||||

QY 181 ATGCTTGACACATGGGATCAGACCTTACCTTACCCCATGAGGCTTGCTT 231
|||||
Db -32844 ATGCTTGACACATGGGATCAGACCTTACCTTACCCCATGAGGCTTGCTT 32894
|||||

- RESULT 6
AC100424/c
LOCUS Mus musculus clone RP23-136F8, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC100424
ACCESSION AC100424
VERSION AC100424.1 GI:17047790
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 66393)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-136F8
Unpublished
2 (bases 1 to 66393)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Kartas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Royle,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L15312
 Center clone name: 136_F_8

* NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 692: contig of 692 bp in length
 * 693 792: gap of 100 bp
 * 793 1509: contig of 717 bp in length
 * 1510 1609: gap of 100 bp
 * 1610 2315: contig of 706 bp in length
 * 2316 2415: gap of 100 bp
 * 2416 3132: contig of 717 bp in length
 * 3133 3232: gap of 100 bp
 * 3233 3913: contig of 680 bp in length
 * 3913 4012: gap of 100 bp
 * 4013 4682: contig of 670 bp in length
 * 4683 4782: gap of 100 bp
 * 4783 5477: contig of 695 bp in length
 * 5478 5577: gap of 100 bp
 * 5578 6280: contig of 703 bp in length
 * 6281 6380: gap of 100 bp
 * 6381 7066: contig of 686 bp in length
 * 7067 7166: gap of 100 bp
 * 7167 7870: contig of 704 bp in length
 * 7871 7970: gap of 100 bp
 * 7971 8665: contig of 695 bp in length
 * 8666 8765: gap of 100 bp
 * 8766 9472: contig of 707 bp in length
 * 9473 9572: gap of 100 bp
 * 9573 10230: contig of 658 bp in length
 * 10231 10330: gap of 100 bp
 * 10331 11024: contig of 694 bp in length
 * 11025 11242: gap of 100 bp
 * 1125 11809: contig of 685 bp in length
 * 11810 11909: gap of 100 bp
 * 11910 12589: contig of 680 bp in length
 * 12590 12689: gap of 100 bp
 * 12690 13374: contig of 685 bp in length
 * 13375 13474: gap of 100 bp
 * 13475 14170: contig of 696 bp in length
 * 14171 14270: gap of 100 bp
 * 14271 14967: contig of 697 bp in length
 * 14968 15067: gap of 100 bp
 * 15068 15770: contig of 703 bp in length
 * 15771 15870: gap of 100 bp
 * 15871 16588: contig of 718 bp in length
 * 16589 16688: gap of 100 bp
 * 16689 17357: contig of 669 bp in length
 * 17358 17457: gap of 100 bp
 * 17458 18160: contig of 703 bp in length
 * 18161 18260: gap of 100 bp
 * 18261 18972: contig of 712 bp in length
 * 18973 19072: gap of 100 bp
 * 19073 19724: contig of 652 bp in length
 * 19725 19824: gap of 100 bp
 * 19825 20520: contig of 696 bp in length
 * 20521 20620: gap of 100 bp

* 20621 21326: contig of 706 bp in length
* 21327 21426: gap of 100 bp
* 22427 22117: contig of 691 bp in length
* 22118 22904: gap of 100 bp
* 22218 22904: contig of 687 bp in length
* 22905 23004: gap of 100 bp
* 23005 23705: contig of 701 bp in length
* 23706 23805: gap of 100 bp
* 23806 24497: contig of 692 bp in length
* 24498 24597: gap of 100 bp
* 24598 25272: contig of 675 bp in length
* 25273 25372: gap of 100 bp
* 25373 26083: contig of 711 bp in length
* 26084 26183: gap of 100 bp
* 26184 26860: contig of 677 bp in length
* 26861 26960: gap of 100 bp
* 26961 27628: contig of 668 bp in length
* 27629 27728: gap of 100 bp
* 27729 28417: contig of 689 bp in length
* 28418 28517: gap of 100 bp
* 28518 29199: contig of 682 bp in length
* 29200 29299: gap of 100 bp
* 29300 29557: contig of 658 bp in length
* 29558 30057: gap of 100 bp
* 30058 30742: contig of 685 bp in length
* 30743 30842: gap of 100 bp
* 30843 31537: contig of 695 bp in length
* 31538 31637: gap of 100 bp
* 31638 32348: contig of 711 bp in length
* 32349 32448: gap of 100 bp
* 32449 33128: contig of 680 bp in length
* 33129 33228: gap of 100 bp
* 33229 33910: contig of 682 bp in length
* 33911 34010: gap of 100 bp
* 34011 34723: contig of 713 bp in length
* 34724 34823: gap of 100 bp
* 34824 35524: contig of 701 bp in length
* 35525 35624: gap of 100 bp
* 35625 36304: contig of 680 bp in length
* 36305 36404: gap of 100 bp
* 36405 37092: contig of 688 bp in length
* 37093 37192: gap of 100 bp
* 37193 37879: contig of 687 bp in length
* 37880 37979: gap of 100 bp
* 37980 38665: contig of 686 bp in length
* 38666 38765: gap of 100 bp
* 38766 39434: contig of 669 bp in length
* 39435 39534: gap of 100 bp
* 39535 40251: contig of 717 bp in length
* 40252 40351: gap of 100 bp
* 40352 41047: contig of 696 bp in length
* 41048 41147: gap of 100 bp
* 41148 41821: contig of 674 bp in length
* 41822 41921: gap of 100 bp
* 41922 42631: contig of 710 bp in length
* 42632 42731: gap of 100 bp
* 42732 43437: contig of 706 bp in length
* 43438 43537: gap of 100 bp
* 43538 44243: contig of 706 bp in length
* 44244 44343: gap of 100 bp
* 44344 45011: contig of 668 bp in length
* 45012 45111: gap of 100 bp
* 45112 45789: contig of 678 bp in length
* 45790 45889: gap of 100 bp
* 45890 46583: contig of 694 bp in length
* 46584 46683: gap of 100 bp
* 46684 47363: contig of 680 bp in length
* 47364 47463: gap of 100 bp
* 47464 48166: contig of 703 bp in length
* 48167 48266: gap of 100 bp
* 48267 48971: contig of 705 bp in length
* 48972 49071: gap of 100 bp
* 49072 49774: contig of 703 bp in length

* 49775 49874: gap of 100 bp
* 49875 50562: contig of 688 bp in length
* 50563 50662: gap of 100 bp
* 50663 51336: contig of 674 bp in length
* 51337 51436: gap of 100 bp
* 51437 52136: contig of 700 bp in length
* 52137 52236: gap of 100 bp
* 52237 52967: contig of 731 bp in length
* 52968 53067: gap of 100 bp
* 53068 53748: contig of 681 bp in length
* 53749 53848: gap of 100 bp
* 53849 54539: contig of 691 bp in length
* 54540 54639: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 66393;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTTCATCCTGTGGAAGG 60
DB 22021 GTGGGTGCGAGGCTAAGCACTGCACAGAGGATAGCTTGTTCATCCTGTGGAAGG 21962
QY 61 CACGCTGATTCGATGAAGGTTCAAGTCTCCTAGTTCCTCCCTCCAGGAAACACACACG 120
DB 21961 CACGCTGATTCGATGAAGGTTCAAGTCTCCTAGTTCCTCCCTCCAGGAAACACACACG 21902
QY 121 GGAGCTGGCCAAACACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 180
DB 21901 GGAGCTGGCCAAACACCTCTCTGGGTGATGAGCCTAAGGGATGTTTGTGTAGGGCCCT 21842
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
DB 21841 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 21791

RESULT 7
AC101278 66489 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-101P16, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC101278
ACCESSION AC101278.1 GI:17060053
VERSION
KEYWORDS HTG: HTGS, PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE 1 (bases 1 to 66489)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-101P16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 66489)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fargo, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L16354
Center clone name: 10L_P_16

* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
723: contig of 723 bp in length
724 823: gap of 100 bp
824 1542: contig of 719 bp in length
1543 1642: gap of 100 bp
1643 2341: contig of 699 bp in length
2342 2441: gap of 100 bp
2442 3174: contig of 733 bp in length
3175 3274: gap of 100 bp
3275 4014: contig of 740 bp in length
4015 4114: gap of 100 bp
4115 4846: contig of 732 bp in length
4847 4946: gap of 100 bp
4947 5656: contig of 710 bp in length
5657 5756: gap of 100 bp
5757 6473: contig of 717 bp in length
6474 6573: gap of 100 bp
6574 7317: contig of 744 bp in length
7318 7417: gap of 100 bp
7419 8129: contig of 712 bp in length
8130 8229: gap of 100 bp
8230 8953: contig of 724 bp in length
8954 9053: gap of 100 bp
9054 9838: contig of 785 bp in length
9839 9939: gap of 100 bp
9939 10667: contig of 729 bp in length
10668 10767: gap of 100 bp
10768 11503: contig of 736 bp in length
11504 11603: gap of 100 bp
11604 12346: contig of 743 bp in length
12347 12446: gap of 100 bp
12447 13168: contig of 722 bp in length
13169 13268: gap of 100 bp
13269 13995: contig of 727 bp in length
13996 14095: gap of 100 bp
14096 14810: contig of 715 bp in length
14811 14910: gap of 100 bp
14911 15587: contig of 677 bp in length
15588 16390: contig of 703 bp in length
16391 16490: gap of 100 bp
16491 17212: contig of 722 bp in length
17213 17312: gap of 100 bp
17313 18033: contig of 721 bp in length
18034 18133: gap of 100 bp
18134 18856: contig of 723 bp in length
18857 18956: gap of 100 bp
18957 19693: contig of 737 bp in length
19694 19794: gap of 100 bp
19794 20529: contig of 736 bp in length
20530 20629: gap of 100 bp
20630 21364: contig of 735 bp in length
21365 21464: gap of 100 bp
21465 22189: contig of 725 bp in length
22190 22289: gap of 100 bp
22290 22981: contig of 692 bp in length
22982 23081: gap of 100 bp
23082 23781: contig of 699 bp in length
23781 23880: gap of 100 bp
23881 24606: contig of 726 bp in length
24607 24707: gap of 100 bp
24707 25424: contig of 718 bp in length
25425 25525: gap of 100 bp
25526 26243: contig of 719 bp in length
26244 26344: gap of 100 bp
26345 27068: contig of 725 bp in length
27069 27168: gap of 100 bp
27169 27900: contig of 732 bp in length
27901 28000: gap of 100 bp
28001 28726: contig of 726 bp in length
28727 28826: gap of 100 bp
28827 29559: contig of 733 bp in length
29560 29659: gap of 100 bp
29660 30397: contig of 738 bp in length
30398 31212: contig of 715 bp in length
31213 31313: gap of 100 bp
31314 32025: contig of 713 bp in length
32026 32125: gap of 100 bp
32126 32855: contig of 730 bp in length
32856 32955: gap of 100 bp
32956 33674: contig of 719 bp in length
33675 33774: gap of 100 bp
33775 34480: contig of 706 bp in length
34481 34580: gap of 100 bp
34581 35294: contig of 714 bp in length
35295 35394: gap of 100 bp
35395 36121: contig of 727 bp in length
36122 36221: gap of 100 bp
36222 36952: contig of 731 bp in length
36953 37052: gap of 100 bp
37053 37771: contig of 719 bp in length
37772 37871: gap of 100 bp
37872 38601: contig of 730 bp in length
38602 38701: gap of 100 bp
38702 39431: contig of 730 bp in length
39432 39531: gap of 100 bp
39532 40256: contig of 725 bp in length
40257 40357: gap of 100 bp
40357 41067: contig of 711 bp in length
41068 41167: gap of 100 bp
41168 41897: contig of 730 bp in length
41898 41997: gap of 100 bp
41998 42712: contig of 715 bp in length
42713 42812: gap of 100 bp
42813 43533: contig of 721 bp in length
43534 43633: gap of 100 bp
43634 44454: contig of 721 bp in length
44455 44555: gap of 100 bp
44556 45187: contig of 733 bp in length
45188 45287: gap of 100 bp
45288 46026: contig of 739 bp in length
46027 46126: gap of 100 bp
46127 46824: contig of 698 bp in length
46825 46924: gap of 100 bp
46925 47656: contig of 732 bp in length
47657 47756: gap of 100 bp
47757 48484: contig of 728 bp in length
48485 48584: gap of 100 bp
48585 49301: contig of 717 bp in length
49302 49401: gap of 100 bp

```

* 49402 50119: contig of 718 bp in length
* 50120 50219: gap of 100 bp
* 50220 50934: contig of 715 bp in length
* 50935 51034: gap of 100 bp
* 51035 51766: contig of 732 bp in length
* 51767 51866: gap of 100 bp
* 51867 52575: contig of 709 bp in length
* 52576 52675: gap of 100 bp
* 52676 53400: contig of 725 bp in length
* 53401 53500: gap of 100 bp
* 53501 54217: contig of 717 bp in length
* 54218 54317: gap of 100 bp
* 54318 55003: contig of 686 bp in length
* 55004 55103: gap of 100 bp
* 55104 55811: contig of 708 bp in length
* 55812 55911: gap of 100 bp
* 55912 56634: contig of 723 bp in length
* 56635 56734: gap of 100 bp

Query Match      100.0%; Score 231; DB 2; Length 66489;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTGTGTCGTCATCCTGTGAAGG 60
Db 19918 GTGGGTGGAGGCTAAGCACTGCACAGAGGATAGCTGTGTCGTCATCCTGTGAAGG 19977
QY 61 CAGCTGTGATGCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAACGACACG 120
Db 19978 CAGCTGTGATGCATGAAGGTTTCAGTGTCTTCCCTTCCCTCCAGGAAACGACACG 20037
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 180
Db 20038 GGAGCTGGCCAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTTGTAGGGCCCT 20097
QY 181 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
Db 20098 ATGCTTGCACACTGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 20148

RESULT 8
AC101278/c
LOCUS
DEFINITION Mus musculus clone RP23-101P16, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC101278
VERSION 1 GI:17060053
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 66489)
Birren, B., Linton, L., Musbaum, C., and Lander, E.
Mus musculus, clone RP23-101P16
Unpublished
2 (bases 1 to 66489)
Birren, B., Linton, L., Musbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,

```

TITLE
JOURNAL
COMMENT

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L16354

Center clone name: 101_P_16

* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
723: contig of 723 bp in length
724 823: gap of 100 bp
824 1542: contig of 719 bp in length
1543 1642: gap of 100 bp
1643 2341: contig of 699 bp in length
2342 2441: gap of 100 bp
2442 3174: contig of 733 bp in length
3175 3274: gap of 100 bp
3275 4014: contig of 740 bp in length
4015 4114: gap of 100 bp
4115 4846: contig of 732 bp in length
4847 4946: gap of 100 bp
4947 5657: contig of 710 bp in length
5657 6474: gap of 100 bp
6474 6573: contig of 717 bp in length
6574 7317: contig of 744 bp in length
7318 7417: gap of 100 bp
7418 8129: contig of 712 bp in length
8130 8229: gap of 100 bp
8230 8953: contig of 724 bp in length
8954 9054: gap of 100 bp
9054 9838: contig of 785 bp in length
9839 10667: gap of 100 bp
10667 10767: contig of 729 bp in length
10768 11503: contig of 736 bp in length
11504 11603: gap of 100 bp
11604 12346: contig of 743 bp in length
12347 12446: gap of 100 bp
12447 13168: contig of 722 bp in length
13169 13268: gap of 100 bp
13269 13995: contig of 727 bp in length
13996 14095: gap of 100 bp
14096 14810: contig of 715 bp in length
14811 14910: gap of 100 bp
14911 15587: contig of 677 bp in length
15588 16390: gap of 100 bp
16391 16490: contig of 703 bp in length
16491 17212: gap of 100 bp
17213 17312: contig of 722 bp in length
17313 17312: gap of 100 bp

* 17313 18033: contig of 721 bp in length
* 18034 18133: gap of 100 bp
* 18134 18856: contig of 723 bp in length
* 18857 18956: gap of 100 bp
* 18957 19693: contig of 737 bp in length
* 19694 19793: gap of 100 bp
* 19794 20529: contig of 736 bp in length
* 20530 20629: gap of 100 bp
* 20630 21364: contig of 735 bp in length
* 21365 21464: gap of 100 bp
* 21465 22189: contig of 725 bp in length
* 22190 22899: gap of 100 bp
* 22899 22981: contig of 692 bp in length
* 22982 23081: gap of 100 bp
* 23082 23780: contig of 699 bp in length
* 23781 23880: gap of 100 bp
* 23881 24606: contig of 726 bp in length
* 24607 24706: gap of 100 bp
* 24707 25424: contig of 718 bp in length
* 25425 25524: gap of 100 bp
* 25525 26243: contig of 719 bp in length
* 26244 26343: gap of 100 bp
* 26344 27068: contig of 725 bp in length
* 27069 27168: gap of 100 bp
* 27169 27900: contig of 732 bp in length
* 27901 28000: gap of 100 bp
* 28001 28726: contig of 726 bp in length
* 28727 28826: gap of 100 bp
* 28827 29559: contig of 733 bp in length
* 29560 29659: gap of 100 bp
* 29660 30397: contig of 738 bp in length
* 30398 30497: gap of 100 bp
* 30498 31212: contig of 715 bp in length
* 31213 31312: gap of 100 bp
* 31313 32025: contig of 713 bp in length
* 32026 32125: gap of 100 bp
* 32126 32855: contig of 730 bp in length
* 32856 32956: gap of 100 bp
* 32957 33674: contig of 719 bp in length
* 33675 34480: contig of 706 bp in length
* 34481 34580: gap of 100 bp
* 34581 35294: contig of 714 bp in length
* 35295 35394: gap of 100 bp
* 35395 36121: contig of 727 bp in length
* 36122 36221: gap of 100 bp
* 36222 36952: contig of 731 bp in length
* 36953 37052: gap of 100 bp
* 37053 37771: contig of 719 bp in length
* 37772 37871: gap of 100 bp
* 37872 38601: contig of 730 bp in length
* 38602 38701: gap of 100 bp
* 38702 39431: contig of 730 bp in length
* 39432 39531: gap of 100 bp
* 39532 40257: contig of 725 bp in length
* 40257 40356: gap of 100 bp
* 40357 41067: contig of 711 bp in length
* 41068 41167: gap of 100 bp
* 41168 41897: contig of 730 bp in length
* 41898 41997: gap of 100 bp
* 41998 42712: contig of 715 bp in length
* 42713 42812: gap of 100 bp
* 42813 43533: contig of 721 bp in length
* 43534 43633: gap of 100 bp
* 43634 44354: contig of 721 bp in length
* 44355 44455: gap of 100 bp
* 44455 45187: contig of 733 bp in length
* 45188 45287: gap of 100 bp
* 45288 46026: contig of 739 bp in length
* 46027 46126: gap of 100 bp
* 46127 46824: contig of 698 bp in length
* 46825 46924: gap of 100 bp
* 46925 47656: contig of 732 bp in length

* 47657 47756: gap of 100 bp
* 47757 48484: contig of 728 bp in length
* 48485 48584: gap of 100 bp
* 48585 49301: contig of 717 bp in length
* 49302 49401: gap of 100 bp
* 49402 50119: contig of 718 bp in length
* 50120 50219: gap of 100 bp
* 50220 50934: contig of 715 bp in length
* 50935 51034: gap of 100 bp
* 51035 51766: contig of 732 bp in length
* 51767 51866: gap of 100 bp
* 51867 52575: contig of 709 bp in length
* 52576 52676: gap of 100 bp
* 52677 53400: contig of 725 bp in length
* 53401 54217: contig of 717 bp in length
* 54218 54317: gap of 100 bp
* 54318 55003: contig of 686 bp in length
* 55004 55103: gap of 100 bp
* 55104 55811: contig of 708 bp in length
* 55812 55911: gap of 100 bp
* 55912 56634: contig of 723 bp in length
* 56635 56734: gap of 100 bp

Query Match 100.0%; Score 231; DB 2; Length 66489;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGTCCGAGGCTAAGCACTGCACAGAGATAGCTGCTGTTGGCATCTCTGGAAGG 60
Db 44342 GTGGGTCCGAGGCTAAGCACTGCACAGAGATAGCTGCTGTTGGCATCTCTGGAAGG 44283

Qy 61 CACGCTGATTCATGAAGTTCAGTCTCCTAGTTCCTCCCTCCAGCAAGCAAGCAAGC 120
Db 44282 CACGCTGATTCATGAAGTTCAGTCTCCTAGTTCCTCCCTCCAGCAAGCAAGC 44223

Qy 121 GGAGCTGGCAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTGTGTAGGGCCCT 180
Db 44222 GGAGCTGGCAAGACCTCTCTGGGTGATGAGCCTTAAGGATGTTGTGTAGGGCCCT 44163

Qy 181 ATGCTTCACACTGGGATCAGACCTCTACCTTACCCTCAGGAGCTTGCCTT 231
Db 44162 ATGCTTCACACTGGGATCAGACCTCTACCTTACCCTCAGGAGCTTGCCTT 44112

RESULT 9
AC101122/c 70360 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP23-93E18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC101122
ACCESSION AC101122
VERSION AC101122.1 GI:17059896
KEYWORDS HTG; HTGS, PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 70360)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
TITLE Mus musculus, clone RP23-93E18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 70360)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faroo, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,

Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Navlor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Reil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: LI4988

Center clone name: 93_E_18

* NOTE: This record contains 86 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 721: contig of 721 bp in length
 722 821: gap of 100 bp
 822 1528: contig of 707 bp in length
 1529 1628: gap of 100 bp
 1629 2332: contig of 704 bp in length
 2333 2432: gap of 100 bp
 2433 3144: contig of 712 bp in length
 3145 3244: gap of 100 bp
 3245 3967: contig of 723 bp in length
 3968 4067: gap of 100 bp
 4068 4791: contig of 724 bp in length
 4792 4891: gap of 100 bp
 4892 5599: contig of 708 bp in length
 5600 5699: gap of 100 bp
 5700 6425: contig of 726 bp in length
 6426 6525: gap of 100 bp
 6526 7219: contig of 694 bp in length
 7220 7319: gap of 100 bp
 7320 8037: contig of 718 bp in length
 8038 8137: gap of 100 bp
 8138 8846: contig of 709 bp in length
 8847 8946: gap of 100 bp
 8947 9664: contig of 718 bp in length
 9665 9764: gap of 100 bp
 9765 10491: contig of 727 bp in length
 10492 10591: gap of 100 bp
 10592 11316: contig of 725 bp in length
 11317 11416: gap of 100 bp
 11417 12126: contig of 710 bp in length
 12127 12226: gap of 100 bp
 12227 12943: contig of 717 bp in length
 12944 13043: gap of 100 bp
 13044 13755: contig of 712 bp in length
 13756 13855: gap of 100 bp
 13856 14559: contig of 704 bp in length
 14560 14659: gap of 100 bp
 14660 15404: contig of 745 bp in length

15405 15504: gap of 100 bp
 15505 16228: contig of 724 bp in length
 16229 16328: gap of 100 bp
 16329 17033: contig of 705 bp in length
 17034 17133: gap of 100 bp
 17134 17838: contig of 705 bp in length
 17839 17938: gap of 100 bp
 17939 18664: contig of 726 bp in length
 18665 18764: gap of 100 bp
 18765 19475: contig of 711 bp in length
 19476 19575: gap of 100 bp
 19576 20305: contig of 730 bp in length
 20306 20405: gap of 100 bp
 20406 21141: contig of 736 bp in length
 21142 21241: gap of 100 bp
 21242 21965: contig of 724 bp in length
 21966 22804: gap of 100 bp
 22805 22804: contig of 739 bp in length
 22806 23637: contig of 733 bp in length
 23638 23737: gap of 100 bp
 23738 24472: contig of 735 bp in length
 24473 24572: gap of 100 bp
 24573 25300: contig of 728 bp in length
 25301 25400: gap of 100 bp
 25401 26122: contig of 722 bp in length
 26123 26222: gap of 100 bp
 26223 26935: contig of 713 bp in length
 26936 27035: gap of 100 bp
 27036 27770: contig of 735 bp in length
 27771 27870: gap of 100 bp
 27872 28550: contig of 680 bp in length
 28551 28650: gap of 100 bp
 28651 29358: contig of 708 bp in length
 29359 29458: gap of 100 bp
 29459 30183: contig of 725 bp in length
 30184 30283: gap of 100 bp
 30284 31012: contig of 729 bp in length
 31013 31112: gap of 100 bp
 31113 31829: contig of 717 bp in length
 31830 31929: gap of 100 bp
 31930 32642: contig of 713 bp in length
 32643 32742: gap of 100 bp
 32743 33463: contig of 721 bp in length
 33464 33563: gap of 100 bp
 33564 34280: contig of 717 bp in length
 34281 34380: gap of 100 bp
 34381 35089: contig of 709 bp in length
 35090 35189: gap of 100 bp
 35190 35922: contig of 733 bp in length
 35923 36022: gap of 100 bp
 36023 36739: contig of 717 bp in length
 36740 36839: gap of 100 bp
 36840 37552: contig of 713 bp in length
 37553 37652: gap of 100 bp
 37653 38375: contig of 723 bp in length
 38376 38475: gap of 100 bp
 38476 39195: contig of 720 bp in length
 39196 39295: gap of 100 bp
 39296 40019: contig of 724 bp in length
 40020 40119: gap of 100 bp
 40120 40844: contig of 725 bp in length
 40845 40944: gap of 100 bp
 40945 41660: contig of 716 bp in length
 41661 41760: gap of 100 bp
 41761 42445: contig of 685 bp in length
 42446 42545: gap of 100 bp
 42546 43241: contig of 696 bp in length
 43242 43341: gap of 100 bp
 43342 44062: contig of 721 bp in length
 44063 44162: gap of 100 bp
 44163 44878: contig of 716 bp in length
 44879 44978: gap of 100 bp

TITLE JOURNAL

COMMENT

1 (bases 1 to 77127)
Wallis, J.
Direct Submission
Submitted (17-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 18, 2003 this sequence version replaced gi:29500502.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep RP24-129F5 is
from a Male (C57BL/6J) mouse BAC Library VECTOR: pTARBAC1.

FEATURES
source
1. 77127
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP24-129F5"
/clone_lib="RPCI-24"
BASE COUNT 21919 a 15230 c 15167 g 24811 t
ORIGIN
Query Match 100.0%; Score 231; DB 10; Length 77127;
Best Local Similarity 100.0%; Pred. No. 1.7e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGTGGAGCGCTAAGCACTGCACAGATAGCTGCTGGCATCTCTGGGAAGG 60
|||||
Db GTGGGGTGGAGCGCTAAGCACTGCACAGATAGCTGCTGGCATCTCTGGGAAGG 33237
QY 61 CAGCTGTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCCAGAAACGACACG 120
|||||
Db CAGCTGTGATTCATGAAGGTTTCAGTGTCTAGTTCCTTCCCCCAGAAACGACACG 33267
QY 121 GGAGCTGGCCAGACCTCTCTGGGTGATGACGCTTAAGGGATGTTTGTGTAGGCCCT 180
|||||
Db GGAGCTGGCCAGACCTCTCTGGGTGATGACGCTTAAGGGATGTTTGTGTAGGCCCT 32207
QY 181 ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGCTTGCTT 231
|||||
Db ATGCTTGACACATGGGGATCAGACCTCTACCTTCACCCATGAGCTTGCTT 32156
RESULT 12
AF481949/c 84250 bp DNA linear ROD 23-OCT-2002
LOCUS Mus musculus DOM1 (Dom1) and DOM2 (Dom2) genes, exons 1 through 5
DEFINITION

```

exon      65543..65571
/gene="Dom2"
/number=1
exon      71054..71685
/gene="Dom2"
/number=2
CDS       join(71058..71685,73128..73398,74282..74432,75229..75420)
/gene="Dom2"
/codon_start=1
/product="Dom2"
/protein_id="AAM47489.1"
/db_xref="GI:21322148"
/translation="MTPSISWGLLLAGLCLVPSELFARDVQETDSQKQSPASHEI
ATNGLFALSKRELVLHQNTNIFPVSFIATAFAMLSIGSKGDTHTQILBGLQFNL
TGTSEADHISQHLQTLNRDPSQLSTGSLFVNNDLKLVKFLPEAKHYQAEV
FVSNFAESEAKKVINDFEKGTQKIVAEKVELDQDTFALANKILFKGKWKPFDP
ENTFAEFHVDKSTVTKVPMVTLGMLDVHSGTSLSSVLLMDYVGNASAVFLPEDG
KMOHLQTLSSKELSKILLNRHRLVQIHIPLRISIGDYNLKLMSPLGITRIFNNGA
DLSGITEENAPLKLKAVHKAVLTIDETGTEAANAATVFAVPMSPPIILRDPHPLFI
IFEHTQSPFVKGKVVDPTHK"
/number=3
/number=4
/number=5
BASE COUNT 2271 a 19868 c 19190 g 22921 t
ORIGIN
```

```

Query Match      100.0%; Score 231; DB 10; Length 84250;
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGCTGCGAGCTAAGCACTGCACAGAGGATGCTTGTGGCATCCTGTGGAAG 60
DB 8605 GTGGGCTGCGAGCTAAGCACTGCACAGAGGATGCTTGTGGCATCCTGTGGAAG 8546

QY 61 CAGCTCTGATTGATGAAGCTTCAAGTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACAG 120
DB 8545 CAGCTCTGATTGATGAAGCTTCAAGTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACAG 8486

QY 121 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGCTTTGTGTAGGCCCCCT 180
DB 8485 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGCTTTGTGTAGGCCCCCT 8426

QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
DB 8425 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 8375
```

```

RESULT 13
AL731664 98653 bp DNA linear ROD 24-JUN-2002
LOCUS Mouse DNA sequence from clone RP23-400M24 on chromosome 4, complete
DEFINITION sequence.
ACCESSION AL731664
VERSION AL731664
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 98653)
JOURNAL Lovell, J.
Direct Submission
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 26, 2002 this sequence version replaced gi:21540117.
----- Genome Center
```

```

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6.
```

FEATURES

```

Source
1..98653
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosomes="4"
/clone="RP23-400M24"
/clone_lib="RPCI-23"
BASE COUNT 28064 a 22581 c 22820 g 25188 t
ORIGIN
```

```

Query Match      100.0%; Score 231; DB 10; Length 98653;
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGGTGCGAGCTAAGCACTGCACAGAGGATGCTTGTGGCATCCTGTGGAAG 60
DB 43322 GTGGGTGCGAGCTAAGCACTGCACAGAGGATGCTTGTGGCATCCTGTGGAAG 43381

QY 61 CAGCTCTGATTGATGAAGCTTCAAGTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACAG 120
DB 43382 CAGCTCTGATTGATGAAGCTTCAAGTTCAGTGTCTTCCCTTCCCTCCAGGAAACACACAG 43441

QY 121 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGCTTTGTGTAGGCCCCCT 180
DB 43442 GGAGCTGGCCAAACCTCTCTGGGTGATGAGCCTAAGGATGCTTTGTGTAGGCCCCCT 43501

QY 181 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 231
DB 43502 ATGCTTGCACATGGGGATCAGACCTCTACCTTCACCCATGAGGCTTGCTT 43552
```

RESULT 14

```

AC132401_0/c
WPCOMMENT
Sequence split into 5 fragments LOCUS AC132401 Accession AC132401
Fragment Name Begin End
AC132401_0 1 110000
AC132401_1 100001 210000
AC132401_2 200001 310000
AC132401_3 300001 410000
AC132401_4 400001 489146
LOCUS AC132401 489146 bp DNA linear HTG 03-SEP-2002
DEFINITION Mus musculus chromosome UNK clone RP23-40A8, WORKING DRAFT
SEQUENCE, 140 unordered pieces.
ACCESSION AC132401
```

```

VERSION      AC132401.1  GI:22657894
KEYWORDS     HTG: HTGS_PHASE1: HTGS_DRAFT.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 489146)
AUTHORS      McPherson,J.D. and Waterston,R.H.
TITLE        The sequence of Mus musculus clone
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 489146)
TITLE        McPherson,J.D. and Waterston,R.H.
JOURNAL      Direct Submission
COMMENT      Submitted (03-SEP-2002) Genome Sequencing Center, 4444 Forest Park
              Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BA0404A08
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Consensus quality: 430636 bases at least Q40
Consensus quality: 448443 bases at least Q30
Consensus quality: 460211 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 140 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1212: contig of 1212 bp in length
1213 1312: gap of unknown length
1313 2777: contig of 1265 bp in length
2578 2677: gap of unknown length
2678 4009: contig of 1732 bp in length
4410 4509: gap of unknown length
4510 5592: contig of 1183 bp in length
5593 5792: gap of unknown length
5793 6999: contig of 1207 bp in length
7000 7099: gap of unknown length
7100 8301: contig of 1202 bp in length
8302 8401: gap of unknown length
8402 9700: contig of 1299 bp in length
9701 9800: gap of unknown length
9801 10916: contig of 1116 bp in length
10917 11016: gap of unknown length
11017 12583: contig of 1567 bp in length
12584 12683: gap of unknown length
12684 13755: contig of 1072 bp in length
13756 13855: gap of unknown length
13856 15575: contig of 1720 bp in length
15576 15676: gap of unknown length
15676 16894: contig of 1219 bp in length
16895 16994: gap of unknown length
16995 18323: contig of 1329 bp in length
18324 18423: gap of unknown length
18424 19659: contig of 1236 bp in length
19660 19759: gap of unknown length
19760 21346: contig of 1587 bp in length
21347 21446: gap of unknown length
21447 22722: contig of 1276 bp in length
22723 22822: contig of 1060 bp in length
22823 23882: gap of unknown length
23883 25115: contig of 1233 bp in length
25116 25316: gap of unknown length
25316 26372: contig of 1057 bp in length
26373 27911: gap of unknown length
27912 28011: gap of unknown length
28012 29838: contig of 1827 bp in length
29839 29938: gap of unknown length
29939 31555: contig of 1617 bp in length
31556 33922: contig of 1537 bp in length
33923 34857: gap of unknown length
34858 36321: contig of 1364 bp in length
36322 38181: contig of 1760 bp in length
38182 39578: gap of unknown length
39579 41309: contig of 1631 bp in length
41310 42539: contig of 1130 bp in length
42540 44944: contig of 2305 bp in length
44945 45045: gap of unknown length
45045 46294: gap of unknown length
46294 47882: contig of 1489 bp in length
47883 49351: contig of 1369 bp in length
49352 50966: contig of 1515 bp in length
50967 52365: contig of 1299 bp in length
52366 54067: contig of 1602 bp in length
54068 55874: gap of unknown length
55875 57285: contig of 1311 bp in length
57286 59624: gap of unknown length
59624 61001: contig of 1278 bp in length
61002 62734: contig of 1633 bp in length
62735 64266: gap of unknown length
64266 66070: contig of 1705 bp in length
66071 68449: contig of 2279 bp in length
68450 69975: gap of unknown length
69975 71869: gap of unknown length
71869 73384: contig of 1416 bp in length
73385 74800: gap of unknown length
74800 77399: contig of 2499 bp in length
77399 78516: contig of 1018 bp in length
78517 80111: contig of 1494 bp in length
80111 81382: contig of 1172 bp in length
81383 81482: gap of unknown length

```

